Factor Analysis of a Large DSGE Model

Alexei Onatski* Francisco J. Ruge-Murcia†

September 27, 2010

Abstract

We study the workings of the factor analysis of high-dimensional data using artificial series generated from a large, multi-sector dynamic stochastic general equilibrium (DSGE) model. The objective is to use the DSGE model as a laboratory that allow us to shed some light on the practical benefits and limitations of using factor analysis techniques on economic data. We explain in what sense the artificial data can be thought of having a factor structure, study the theoretical and finite sample properties of the principal components estimates of the factor space, investigate the substantive reason(s) for the good performance of diffusion index forecasts, and assess the quality of the factor analysis of highly dissagregated data. In all our exercises, we explain the precise relationship between the factors and the basic macroeconomic shocks postulated by the model.

JEL Classification: C3, C5, E3

Keywords: Multisector economies, principal components, forecasting, pervasiveness, FAVAR.

*Faculty of Economics, University of Cambridge, Sidgwick Avenue, Cambridge, CB3 9DD, United Kingdom. E-mail: ao319@econ.cam.ac.uk

†Département de sciences économiques and CIREQ, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal (Québec) H3C 3J7, Canada. E-mail: francisco.ruge-murcia@umontreal.ca

1. Introduction

Large factor models have been playing an increasingly important role in empirical macroeconomic research. Their use is often motivated by the fact that modern dynamic general equilibrium models postulate the existence of only a few common sources of fluctuations for all macroeconomic variables. Much of the research takes this fact a step further and identifies the factor space with the space of basic macroeconomic shocks. This interpretation is then exploited to carry out structural factor analysis.[1] Although the statistical workings of large factor models and of their forecasting and structural applications are well understood by now, their macroeconomic content remains largely unexplored. Usually, the existence of such a macroeconomic content is simply a matter of assumption or believe. In particular, the relationship between the factor space and the space of macroeconomic shocks has never been analyzed in the context of a theoretical macroeconomic model. The reason is that few general equilibrium models describe the dynamics of a large enough number of disaggregated macroeconomic series to provide the basis for a substantive analysis of large factor models.

This paper studies the workings of the factor analysis of high-dimensional data using artificial series generated from a dynamic stochastic general equilibrium (DSGE) model. The intention is to use the fully-specified DSGE model as a laboratory to understand the practical benefits and limitations of using factor analysis techniques on economic data. We pursue this approach because in contrast to actual economic data, whose generating process is unknown, the artificial data from a model comes from a process that is known and under the control of the econometrician. As DSGE model we use the highly disaggregated multi-sector model developed and estimated by Bouakez, Cardia and Ruge-Murcia (2009). Several features of this model make it particularly suitable for our analysis. First, it specifies thirty heterogenous productive sectors that correspond to the two-digit level of the Standard Industrial Classification (SIC) and so it can generate a large number of disaggregated series to which one can meaningfully apply factor analysis techniques. Second, the model features aggregate shocks but also sector-specific shocks which may be transmitted to other sectors through the input-output structure of the economy. Thus, as in the actual data, the notion of what "factors" are is not trivial. Finally, although the model is (by definition) a stylized representation of the economy, it is rich enough to shed some light on the application of factor analysis to actual disaggregated data.

In this project, we explain in what sense the data generated from the model can be thought of as having a factor structure, study the theoretical and finite sample properties of the principal

[1] For two recent examples, see Boivin et al. (2009) and Forni et al. (2009).

components estimates of the factor space, investigate the performance of diffusion index forecasts, and assess the quality of the factor analysis of highly dissagregated data. In all our exercises, we explain the precise relationship between the factors and the basic macroeconomic shocks postulated by the model.

We find that the three economy-wide shocks of Bouakez et al. (2009), namely the monetary policy shock, the money demand shock and the leisure preference shock, can indeed be thought of as factors because they non-trivially affect most of the 156 variables generated by the model. Moreover, the remaining 30 sectoral productivity shocks significantly affect only a small number of the variables. However, despite the pervasiveness of the economy-wide shocks, the principal components analysis has a hard time replicating the macroeconomic factor space. We document and explain the difficulties that arise and assess the quality of the asymptotic approximation to the distribution of the principal components, find it unsatisfactory, and analyze the reasons for such a failure.

Further, we show that the diffusion index forecasts of output growth and of aggregate inflation perform reasonably well on our simulated data. In particular, accurately estimating the macroeconomic factor space turns out to be not essential for the quality of the forecasts of our data. We decompose the diffusion index forecast error into several components and study them in detail.

Finally, we use our data to investigate the workings of the factor augmented vector autoregression (FAVAR) analysis of disaggregated data. We are especially interested in the question of whether the monetary policy impulse responses of the disaggregated series estimated by a FAVAR fitted to our data accurately recover the true impulse responses. Somewhat surprisingly, we find that the quality of the estimated impulse responses of the sectoral variables is very good.

The rest of the paper is organized as follows. Section 2, describes the data generating process. (A very detailed description of the DSGE model and parameter values used to generate the artificial data is also given in the Appendix.) Section 3 explains in what sense our data has a dynamic factor structure and studies the relationship between the space of dynamic factors, represented by the economy-wide shocks, and the space of dynamic principal components. Section 4 compares the spaces spanned by the lags of the dynamic factors and by the population static and generalized principal components. Section 5 studies the determination of the number of factors. Section 6 compares the population and sample principal components. Section 7, analyzes the diffusion index forecasts. Section 8 performs a FAVAR analysis of the monetary policy effects on the disaggregated variables. Finally, Section 9 concludes.

2. Data generating process

The data generating process (DGP) is based on the large multi-sector DSGE model in Bouakez, Cardia and Ruge-Murcia (2009) (BCR in what follows). Their model features thirty sectors or industries that roughly correspond to the two-digit level of the Standard Industrial Classification (SIC). Sectors are heterogenous in production functions, price rigidity, and the combination of materials and investment inputs used to produce their output. The productive structure is of the roundabout form, meaning that each sector uses output from all sectors as inputs, although in a manner consistent with the actual U.S. Input-Output and Capital-Flow Tables. In addition, sectors are subject to idiosyncratic productivity shocks. There is a representative consumer who supplies labor to all sectors, and derives utility from leisure, real money balances, and the consumption of goods produced by all sectors.

Economic fluctuations arise from three aggregate (or economy-wide) shocks and thirty sectoral shocks. The aggregate shocks are shocks to the representative consumer's utility from leisure and from holding real money balances, and a monetary policy shock, which is modeled as a shock to the growth rate of the money supply. The thirty productivity shocks are sector-specific in that they originally disturb only the production function of their own sector. However, as a result of input-output interactions between sectors, idiosyncratic productivity shocks are transmitted to other sectors and to economic aggregates. All the thirty-three shocks of the model are assumed to follow independent univariate AR(1) processes. The model is described in more detail in the Appendix. The values of the model parameters are those estimated by Bouakez, Cardia and Ruge-Murcia (2009) using quarterly aggregate and sectoral U.S. data from 1964:Q2 to 2002:Q4 and are listed in the Appendix as well.

We solve the log-linearized equations of the BCR model using the standard Blanchard and Khan (1980) algorithm. Collecting the log-linear approximations to the equilibrium decision rules and arranging them into the state-space form, we can write the state-space representation of the data generating process:

$$\begin{aligned} X_{t+1} &= AX_t + B\varepsilon_t \\ Y_t &= CX_t + D\varepsilon_t. \end{aligned} \tag{1}$$

The meaning of the variables ε_t, Y_t and X_t is as follows. The 33-dimensional vector ε_t consists of the 30 unit-variance innovations to the sector-specific productivity shocks and three unit variance innovations to the economy-wide shocks, namely the leisure preference shock, the money demand shock and the monetary policy shock. The 156-dimensional vector of simulated data Y_t consists of

stacked 1×30 vectors of the percentage deviations from the steady state of the sectoral outputs, sectoral hours worked, sectoral wages, sectoral consumptions, and sectoral inflations; scalar aggregate output, hours worked, wages, consumption, and inflation; and scalar rates of money growth and nominal interest. The state vector X_t is a linear combination of sectoral and aggregate variables. The choice of the state vector is not unique and it is usually made so that the dimension of the matrix A is as small as possible. In our case, the smallest possible dimension is 101.[2]

For the numerical values of A, B, C and D in (1), we check that the innovations ε_t can be recovered from the history of Y_t, and thus, they are fundamental. Fernandez-Villaverde et al. (2007) describe a simple criterion for checking the fundamentalness of innovations when D is a square invertible matrix. In our case, D is 156×33 so that the criterion cannot be directly applied. However, we can use the criterion for checking the fundamentalness of ε_t in the system

$$\begin{aligned} X_{t+1} &= AX_t + B\varepsilon_t \\ \tilde{Y}_t &= \tilde{C}X_t + \tilde{D}\varepsilon_t, \end{aligned}$$

where $\tilde{Y}_t = R'Y_t$, $\tilde{C} = R'C$, $\tilde{D} = R'D$ and R is any 156×33 matrix. We choose R so that it consists of the first 33 columns of matrix U from the singular value decomposition of $D : D = USV$.[3] Then, we form the matrix $M = A - B\tilde{D}^{-1}\tilde{C}$ and check numerically that all its eigenvalues are less than one in absolute value, which, according to Fernandez-Villaverde et al. (2007), insures that ε_t can be recovered from the history of $\tilde{Y}_t \equiv R'Y_t$, and therefore, from the history of Y_t itself.

Equations (1) can be used to express the simulated data as an infinite lag polynomial of the innovations ε_t :

$$Y_t = \left[CL\left(I - AL\right)^{-1} B + D\right] \varepsilon_t \equiv A(L)\varepsilon_t, \tag{2}$$

where L is the lag operator. The i, j-th entry of the matrix coefficient on the p-th power of L in the polynomial $A(L)$ equals the impulse response of the i-th variable in the data to a one-period unit impulse in the j-th component of the innovation vector ε_t.

3. The dynamic factor structure and principal components

Let f_t be the three-dimensional vector of economy-wide shocks. As was mentioned before, they are modeled as independent AR(1) processes so that

$$f_t = \rho f_{t-1} + \sigma_f \varepsilon_{ft},$$

[2] We used MATLAB's Control System toolbox command *minreal* to obtain the minimal state-space realization in (1), which turned out to have a state vector of dimensionality 101.

[3] Such a choice maximizes the smallest eigenvalue of $\tilde{D}\tilde{D}' = R'DD'R$ among all R with $\|R\| = 1$, where $\|R\|$ denotes the square root of the largest eigenvalue of $R'R$.

where ε_{ft} are the standardized innovations to the economy-wide shocks, σ_f are scaling parameters, and ρ is a 3×3 matrix with the autoregressive coefficients along the main diagonal and zero everywhere else. We can decompose the MA(∞) representation of the model (2) into a part that depends on f_{t-j} with $j = 0, 1, ...$, and the orthogonal part:

$$Y_t = \Lambda(L) f_t + e_t, \tag{3}$$

where $\Lambda(L) = A_f(L) \sigma_f^{-1} (I_3 - \rho L)$ with I_3 the 3×3 identity matrix, $e_t = A_e(L) \varepsilon_{et}$, ε_{et} is the vector of the innovations to the sector-specific shocks, and $A_f(L)$ and $A_e(L)$ are the parts of $A(L)$ corresponding to the economy-wide and sector-specific shocks, respectively.

In what follows, we will always standardize Y_t so that the variance of each of its components equals unity. It is convenient to introduce new notation for such a standardized Y_t. Let W be the inverse of the diagonal matrix with the standard deviations of the components of Y_t on the diagonal. Then, the standardized version of Y_t equals $Y_t^{(s)} = WY_t$, and it admits the following decomposition

$$Y_t^{(s)} = \Lambda^{(s)}(L) f_t + e_t^{(s)},$$

where $\Lambda^{(s)}(L) = W\Lambda(L)$ and $e_t^{(s)} = We_t$.

Intuitively, we would call (3) a factor decomposition if the "factors" f_t had a pervasive effect on the elements of Y_t whereas the "idiosyncratic" terms e_{it} did not have pervasive effects on the elements of Y_t. Such requirements are in the spirit of all high-dimensional factor models starting from the approximate factor model of Chamberlain and Rothschild (1983). We will now examine how pervasive f_t and ε_{et} are in our artificial data.

3.1 Pervasiveness

An intuitive measure of the pervasiveness of a given shock in a particular dataset can be obtained as follows. For each variable in the dataset, compute the percentage of the variance of this variable due to the given shock. Then, compute and plot the percentage $y(z)$ of the variables in the dataset for which the shock explains at least $z\%$ of the variance. Note that such a measure of pervasiveness does not depend on whether we analyze Y_t or its standardized version $Y_t^{(s)}$. Pervasive shocks non-trivially affect a large number of the variables in the dataset. Therefore, for such shocks, $y(z)$ should decrease slowly in z. In contrast, for non-pervasive shocks, we expect $y(z)$ to be dramatically decreasing to zero as z becomes slightly larger than zero.

Figure 1 shows the pervasiveness measures $y(z)$ for all the 33 shocks of the BCR model. As one would intuitively expect, the economy-wide shocks f_t stand out as much more pervasive than the sector-specific shocks. The leisure preference shock is the most pervasive: It explains more than

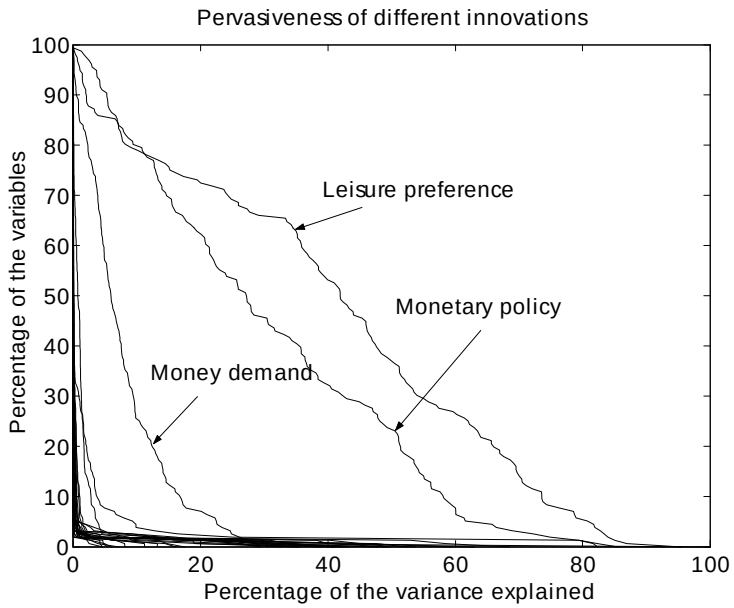

Figure 1: Pervasiveness of different shocks in BCR model.

30% of variance for about 2/3 of all the variables in our dataset. The monetary policy shock is also pervasive, explaining 30% of the variance for more than 40% of the variables. The money demand shock is the least pervasive of the three. Even so, it explains more than 10% of the variance for more than 25% of the variables. The next most pervasive shock is the productivity shock in the agricultural sector: It explains more than 10% of the variance for only about 3.8% of variables.

We can extend our measure of pervasiveness to a frequency-by-frequency measure. In principle, we can define a measure of pervasiveness $y(z, \omega)$ for ε_j as the percentage of the components of Y for which the percentage of the variance at frequency ω explained by ε_j is less than z. However, it would be difficult to report such a pervasiveness measure for the 33 shocks on the same plot. We, therefore, restrict our attention to a pervasiveness measure corresponding to three frequency bands: low frequencies, business cycle frequencies and high frequencies. We define the low frequency band as the set of all frequencies which correspond to cycles of more than 8 years per period, the business cycle frequencies as those corresponding to cycles in between 2 and 8 years per period, and the high frequencies as those corresponding to cycles of less than 2 years per period.

We define $y_j^{LF}(z)$ as the percentage of such $i \in \{i = 1, 2, ..., 156\}$ for which

$$\int_{\omega \in LF} |A_{ij}(\omega)|^2 \, d\omega \geq \frac{z}{100} \sum_{j=1}^{33} \int_{\omega \in LF} |A_{ij}(\omega)|^2 \, d\omega,$$

where $A_{ij}(L)$ is the i, j-th component of $A(L)$ in (2). That is, $y_j^{LF}(z)$ is the percentage of the

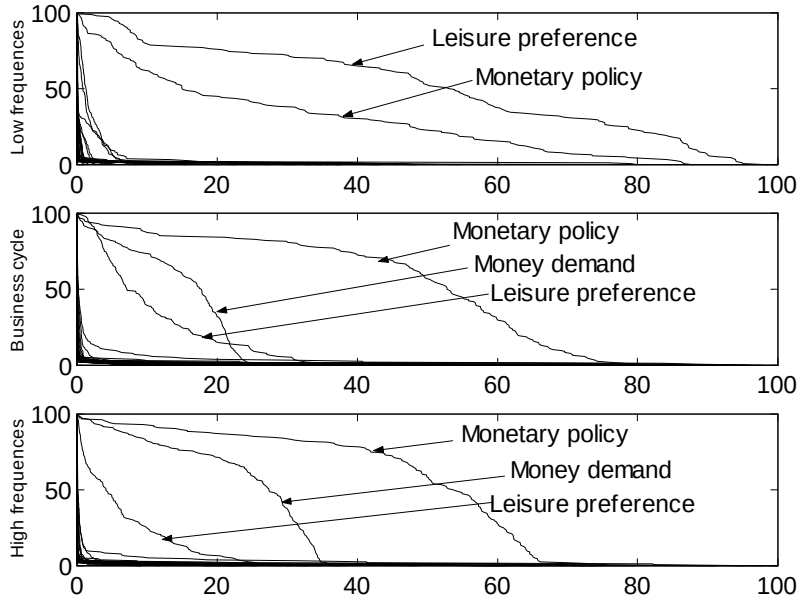

Figure 2: Frequency-by-frequency pervasiveness of different shocks in the BCR model.

variables for which shock j is responsible for at least $z\%$ of the variance at low frequencies. We similarly define $y_j^{BF}(z)$ and $y_j^{HF}(z)$ for business cycle and high frequencies.

Figure 2 shows plots of $y_j^{LF}(z)$ (upper panel), $y_j^{BF}(z)$ (middle panel) and $y_j^{HF}(z)$ (lower panel). We see that for low frequencies only two shocks are pervasive: the leisure preference shock and the monetary policy shock. For business cycle frequencies all three economy-wide shocks are pervasive. However, the leisure preference shock is considerably less pervasive than the money demand shock and, especially, than the monetary policy shock. For high frequencies, the pervasiveness of the leisure preference shock almost completely disappears, whereas the monetary policy and money demand remain pervasive.

Hence, for our data, different frequencies correspond to different shocks being pervasive. Moreover, the number of the pervasive shocks varies from frequency to frequency. This observation suggests that, from the empirical perspective, it may be desirable to relax the assumption of the generalized dynamic factor models that the number of the exploding eigenvalues of the spectral density matrix of the data remains the same for all frequencies.

3.2 Heterogeneity

The pervasiveness of the economy-wide shocks is a necessary but not sufficient condition for successful recovery of the space spanned by all lags and leads of such shocks by the principal components

analysis. For such a recovery, the economy-wide shocks must generate heterogeneous enough responses from the observed variables. The reason is that the model $Y_t^{(s)} = \Lambda^{(s)}(L) f_t + e_t^{(s)}$ is equivalent to $Y_t^{(s)} = \tilde{\Lambda}^{(s)}(L) \tilde{f}_t + e_t^{(s)}$, where $\tilde{\Lambda}^{(s)}(L) = \Lambda^{(s)}(L) U(L)$, $\tilde{f}_t = \overline{U}(L^{-1})' f_t$ and $U(L)$ is a so-called Blaschke matrix (see Lippi and Reichlin, 1994), that is a matrix of polynomials in L such that $\det U(z) \neq 0$ for z on the unit circle and $U(z)\overline{U}(z^{-1})' = I$, where the bar over a polynomial matrix denotes the polynomial matrix with complex conjugated coefficients. Therefore, for the successful recovery, not only f_t, but also $\tilde{f}_t$ must be pervasive. If the columns of $\Lambda(L)$ are not heterogeneous enough so that some of their linear combinations with coefficients that are polynomials in L is small (in terms of the sum of squares of all the coefficients on the different powers of L), then one can choose $U(L)$ so that at least one component of $\tilde{f}_t$ is not pervasive, and therefore, the principal components analysis will not accurately recover the space spanned by all lags and leads of $\tilde{f}_t$, and hence of f_t.

Furthermore, for successful recovery of the space spanned by all lags and leads of the economy-wide shocks by the principal components analysis, the responses from the observed variables to the sector-specific shocks must be heterogeneous enough. Indeed, if such a responses were similar, then there would exist a linear combination of the sector-specific shocks that generate a particularly large response from the observed variables. The variance of the data explained by such a linear combination would be large, and therefore, it could be confused for a genuine factor by the principal components analysis.

A convenient tool for the joint analysis of pervasiveness and heterogeneity is the eigenvalues of the spectral density matrices of the factor and idiosyncratic components of the data. Such eigenvalues are invariant with respect to the different choices of Blaschke matrices $U(L)$ in the above equations. Let $S_f(\omega)$ be the spectral density matrix of $\Lambda^{(s)}(L) f_t$. The largest eigenvalue of $S_f(\omega)$ measures the maximum amount of the variation of the (standardized) data at frequency ω explained by a white-noise shock $\tilde{f}_{1t}$ that belongs to the space spanned by the lags and leads of the economy-wide shocks. The second largest eigenvalue of $S_f(\omega)$ measures the maximum amount of the variation of the data at frequency ω explained by a white-noise shock $\tilde{f}_{2t}$ which is orthogonal to $\tilde{f}_{1t}$ at all lags and leads and which belongs to the space spanned by the lags and leads of the economy-wide shocks. The third largest eigenvalue of $S_f(\omega)$ is the last non-zero eigenvalue of $S_f(\omega)$ because there are only three economy-wide shocks in our model. This eigenvalue measures the minimum amount of the variation of the data at frequency ω explained by a white-noise shock $\tilde{f}_{3t}$ that belongs to the space spanned by the lags and leads of the economy-wide shocks. The three non-zero eigenvalues of $S_f(\omega)$ can be viewed as one-dimensional summaries of the pervasiveness of the shocks $\tilde{f}_{1t}$, $\tilde{f}_{2t}$ and $\tilde{f}_{3t}$. If $\tilde{f}_{3t}$ is not pervasive in the sense that the third largest eigenvalue of

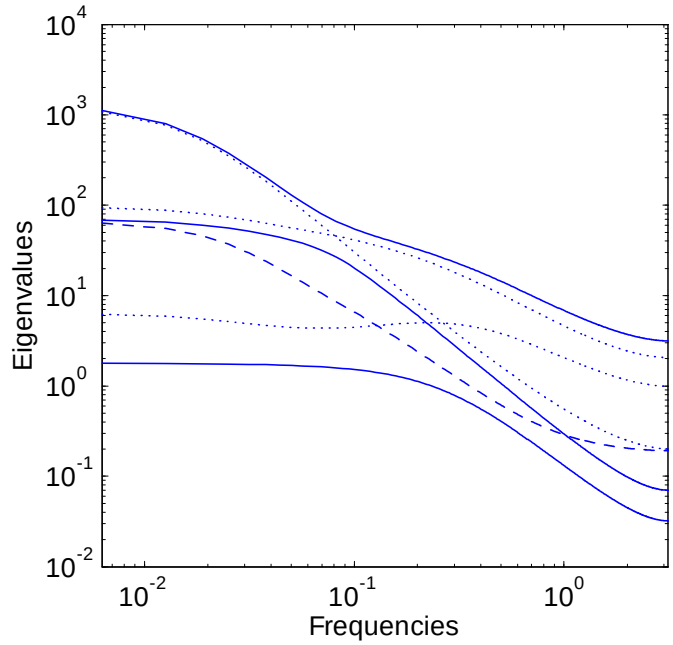

Figure 3: The three largest eigenvalues of $S_f(\omega)$ (solid lines), the largest eigenvalue of $S_e(\omega)$ (dashed line), and the largest eigenvalues of $S_{f1}(\omega)$, $S_{f2}(\omega)$ and $S_{f3}(\omega)$ (dotted lines).

$S_f(\omega)$ is relatively small uniformly over $\omega \in [0, 2\pi)$, then the responses from the observed variables to the economy-shocks f_t must be not heterogeneous. We would hope that the principal components analysis accurately recovers the space spanned by all lags and leads of the economy-wide shocks only if the third largest eigenvalue of $S_f(\omega)$ is larger than the largest eigenvalue of the spectral density matrix $S_e(\omega)$ of the idiosyncratic component $e_t^{(s)}$ uniformly over $\omega \in [0, 2\pi)$.

Figure 3 shows that only the largest eigenvalue of $S_f(\omega)$ is uniformly larger than the largest eigenvalue of $S_e(\omega)$ over $\omega \in [0, 2\pi)$.[4] The second largest eigenvalue of $S_f(\omega)$ is smaller than the largest eigenvalue of $S_e(\omega)$ for relatively high frequencies $\omega > 1$, which correspond to fluctuations with periods smaller than 2π quarters. The third largest eigenvalue of $S_f(\omega)$ is uniformly smaller than the largest eigenvalue of $S_e(\omega)$ over $\omega \in [0, 2\pi)$. Hence, we expect (an imperfect) recovery of at most two orthogonal linear filters of the three-dimensional factor f_t by the principal components analysis.

The dotted lines on Figure 4 show the largest eigenvalues of the spectral densities $S_{f1}(\omega)$, $S_{f2}(\omega)$ and $S_{f3}(\omega)$ of the components of the observables that correspond to the monetary policy shock, to the money demand shock and to the leisure preference shock, respectively. These eigen-

[4] The horizontal and the vertical scales of the graph are made logarithmic to enhance visibility.

values are all larger than the largest eigenvalue of $S_e(\omega)$ at frequencies $\omega > 0.1$, which correspond to fluctuations with periods smaller than 20π quarters, including business cycles, which typically are thought as having periods no larger than 8 or 10 years. Nevertheless, we expect that the business cycle components of the economy-wide shocks cannot be recovered by the principal components analysis. The reason is that the effects of the economy-wide shocks on the observables are not heterogeneous enough, which is reflected in the fact that the third largest eigenvalue of $S_f(\omega)$ is substantially smaller than the minimum of the largest eigenvalues of $S_{f1}(\omega)$, $S_{f2}(\omega)$ and $S_{f3}(\omega)$ for all $\omega \in [0, 2\pi)$.

We checked that the largest eigenvalue of $S_e(\omega)$ is very close to the maximum of the largest eigenvalues of the spectral densities $S_{ei}(\omega)$, $i = 1, ..., 30$ of the thirty components of the data corresponding to the sector-specific shocks. Therefore, the potential problem with the principal components analysis is indeed caused by the similarity in the effects of the economy-wide shocks and not by a possibility that a particular linear filter of the sector-specific shocks have an unusually strong effect on the observables. Similarly, in actual economies different aggregate shocks may also generate similar dynamics on (a subset of) observable variables. For example, in a money-growth targeting regime, changes in monetary aggregates would have similar effects on, say, output variables, regardless of whether they are the result of changes in the monetary base by the central bank or in lending behavior by commercial banks.

3.3 The content of the dynamic principal components

Estimation of the pervasive factors and the factor loadings in large factor models is often based on the sample principal components analysis. In this subsection, we perform the population principal components analysis to determine the theoretical limit to the principal-component-based extraction of the pervasive shocks for our standardized data.

As mentioned above, we expect (an imperfect) recovery of at most two orthogonal linear filters of the three-dimensional factor f_t because the dynamic effects of the economy-wide shocks on the observables are not heterogeneous enough. Specifically, we expect the first two dynamic principal components to be close to some filters of f_t, and the third dynamic principal component to have a very large sector-specific part. We now check this conjecture.

Recall that the dynamic principal components of $Y_t^{(s)}$ are defined as follows.[5] Let $S_Y(\omega)$ be the spectral density matrix of the standardized data and let $\lambda_j(\omega)$ and $p_j(\omega)$ be its j-th largest eigenvalue and the corresponding unit-length row eigenvector, respectively, so that $p_j(\omega) S_Y(\omega) = \lambda_j(\omega) p_j(\omega)$. Consider the Fourier expansion for $p_j(\omega) : p_j(\omega) = (1/2\pi) \sum_{k=-\infty}^{\infty} P_{jk} e^{-ik\omega}$, where

[5] See, also, Brillinger (1981, Ch. 9) and Forni et al. (2000).

$P_{jk} = \int_{-\pi}^{\pi} p_j(\omega) e^{ik\omega} d\omega$. Let $p_j(L)$ equal $(1/2\pi) \sum_{k=-\infty}^{\infty} P_{jk} L^k$. Then, the j-th dynamic principal component of $Y_t^{(s)}$ is defined as $\pi_{jt} \equiv p_j(L) Y_t^{(s)}$. Or, in terms of the innovations to the economic shocks:

$$\pi_{jt} \equiv G_j(L)\varepsilon_t, \tag{4}$$

where $G_j(L) = p_j(L) WA(L)$.

The relationship (4) can be easily and usefully visualized in the frequency domain. Let $\varepsilon_t = \int_{-\pi}^{\pi} e^{i\omega t} dZ_\varepsilon(\omega)$ be the spectral representation of the innovation process ε_t so that $e^{i\omega t} dZ_\varepsilon(\omega)$ can be thought of as the ω-frequency component of ε_t. Then, the spectral representation of π_{jt} has form:

$$\pi_{jt} = \int_{-\pi}^{\pi} e^{i\omega t} G_j(\omega) dZ_\varepsilon(\omega).$$

Hence, the ω-frequency component of π_{jt} is the linear combination of the ω-frequency components of the innovations ε_t with the weights equal to the entries of vector $G_j(\omega) \equiv (G_{j,1}(\omega), ..., G_{j,33}(\omega))$. The functions $(1/2\pi) |G_{j,k}(\omega)|^2$ with $k = 1, ..., 33$ are the spectral densities of the projections of the first dynamic principal component on the spaces spanned by all lags and leads of the innovations $\varepsilon_{t,k}$ with $k = 1, ..., 33$, respectively.

Figure 4 shows the graphs of $(1/2\pi) |G_{1,k}(\omega)|^2$. For some k, the spectral densities are particularly large while for others they are small. We indicate which innovation the dominating densities correspond to. For example, the low frequency component (less than one cycle per 20π quarters, which is about 15 years) of the first dynamic principal component of the data is strongly dominated by the leisure preference innovation. For business cycle frequencies, the monetary policy innovation becomes dominant.[6]

The sum of the areas under the graphs of Figure 4 corresponding to the monetary policy, the money demand and the preference innovations measure the variance of the projection of the first principal component on the space spanned by all lags and leads of the corresponding economy-wide shocks. Such a variance constitutes 97.6% of the overall variance of the first dynamic principal component.

For the second dynamic principal component, the variance of its projection on the space spanned by all lags and leads of the economy-wide shocks is still a respectable 82.6% of its overall variance. However, for the third dynamic principal component, this number is much smaller: 16.8%. In fact, most of the variance of the third dynamic principal component is explained by the innovations to

[6] The logarithmic scale of the horizontal axis of the graph creates the impression that the first dynamic principal component of the data is totally dominated by the leisure preference innovation, which is not the case. The monetary policy innovation explains a comparable portion of the variance of the first dynamic principal component to that explained by the preference innovation.

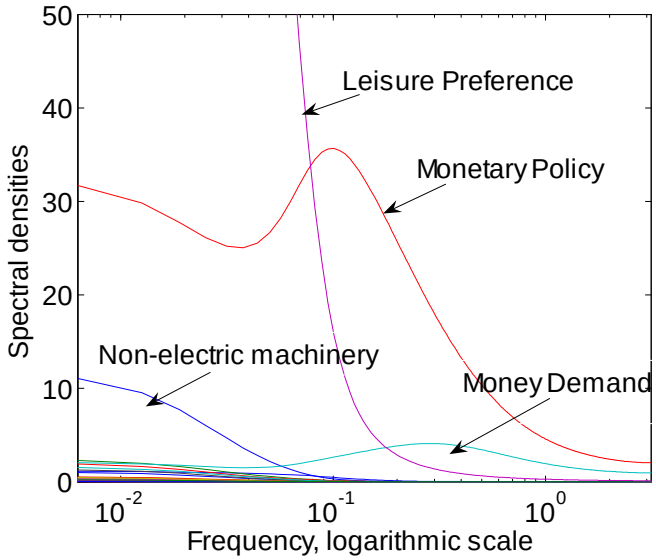

Figure 4: Spectal densities of the projections of the first dynamic principal component on the spaces spanned by all lags and leads of the innovations $\varepsilon_{t,k}$ with $k = 1, ..., 33$.

the agricultural, coal mining and textile mill productivity shocks. The variance of its projection on all lags and leads of these shocks constitutes 67.0% of its overall variance.

Hence, as we had qualitatively expected, the projection of the two first dynamic principal components on the space spanned by all lags and leads of the economy-wide shocks is not much different from the dynamic principal components's themselves. However, the economy-wide content of the third dynamic principal component is very weak. It should be noted here that the first two dynamic principal components explain 80.6% of the variance of the data-generating process while the share of the third dynamic principal component is only 2.4%. Therefore, the agricultural, coal mining and textile mill production shocks do not really have much influence on the economic dynamics as might appear from their substantial share in the dynamics of the third dynamic principal component. The point we would like to stress is simply that the space of the lags and leads of the first three dynamic principal components is substantially different from the space of the lags and leads of the economy-wide shocks f_t.

Can we interpret the first dynamic principal component as a univariate index summarizing the most relevant information in the history of the economy-wide shocks? The answer to this question depends on whether the projection of the first dynamic principal component on the space spanned by leads only of the economy-wide innovations is reasonably small. Let $G_{1,k}(\omega) =$

$(1/2\pi)\sum_{s=-\infty}^{\infty} G_{1,k,s}e^{-is\omega}$ be the Fourier expansion for $G_{1,k}(\omega)$. Let us define

$$\begin{aligned} G_{1,k}^{+}(\omega) &\equiv (1/2\pi)\sum\nolimits_{s=0}^{\infty} G_{1,k,s}e^{-is\omega} \\ G_{1,k}^{-}(\omega) &\equiv (1/2\pi)\sum\nolimits_{s=-\infty}^{-1} G_{1,k,s}e^{-is\omega}. \end{aligned}$$

Then, the variance of the projection of the first dynamic principal component on the leads only of the economy-wide innovations equals

$$(1/2\pi)\int_{-\pi}^{\pi}\left(\left|G_{1,mp}^{-}(\omega)\right|^2+\left|G_{1,md}^{-}(\omega)\right|^2+\left|G_{1,lp}^{-}(\omega)\right|^2\right)d\omega,$$

where $k = mp, md$ and lp for monetary policy, money demand and leisure preference shock innovations, respectively. We have computed this number: It equals 33.5% of the variance in the first dynamic principal component which is due to both the leads and the lags of the economy-wide innovations. The largest part of this percentage (69.9% of it) comes from $\left|f_{1,lp}^{-}(\omega)\right|^2$. That is, the first dynamic principal component has a relatively large projection on the subspace spanned by the future of the leisure preference shock. This finding suggests that in actual applications, the first dynamic principal component may be an imperfect index of the information contained in history of aggregate shocks.

4. Static factors and principal components

In practice, it is often assumed that the dependence of the observables on the factors can be captured by finite lag polynomials. That is, it is assumed that the maximal order of the component polynomials of $\Lambda^{(s)}(L)$ in $Y_t^{(s)} = \Lambda^{(s)}(L) f_t + e_t^{(s)}$ is finite and equal to, say, h. This assumption allows researchers to represent the dynamic factor model in static form by interpreting the h lags of the dynamic factors as additional "static factors". Then, the static principal components or the generalized principal components (see Forni et al., 2005) can be used to recover the static factor space. One advantage of such an approach relative to the dynamic principal components method is that the obtained estimates of the factor space are guaranteed to be orthogonal to the space of the future factor innovations and hence, can be used for forecasting.

The first question we address in this section is whether a few lags of f_t capture most of the information about $Y_t^{(s)}$ contained in $\Lambda^{(s)}(L) f_t$. One way to answer this question is to repeat the above pervasiveness analysis with $\Lambda^{(s)}(L)$ replaced by a finite lag polynomial matrix $\tilde{\Lambda}(L)$. We define $\tilde{\Lambda}(L)$ so that the coefficients of its components $\tilde{\Lambda}_{ij}(L) = \tilde{\Lambda}_{ij,0} + \tilde{\Lambda}_{ij,1}L + ... + \tilde{\Lambda}_{ij,h}L^h$ equal the coefficients of the linear projection of $Y_{it}^{(s)}$ on $f_{jt}, f_{j,t-1}, ...,$ and $f_{j,t-h}$. Figure 5 reports the pervasiveness graph (dashed line) for the leisure preference shocks copied from Figure 1 superimposed with the "finite-lag pervasiveness graphs" for different values of $h = 0, 1, 2, 3$ and 4. Precisely,

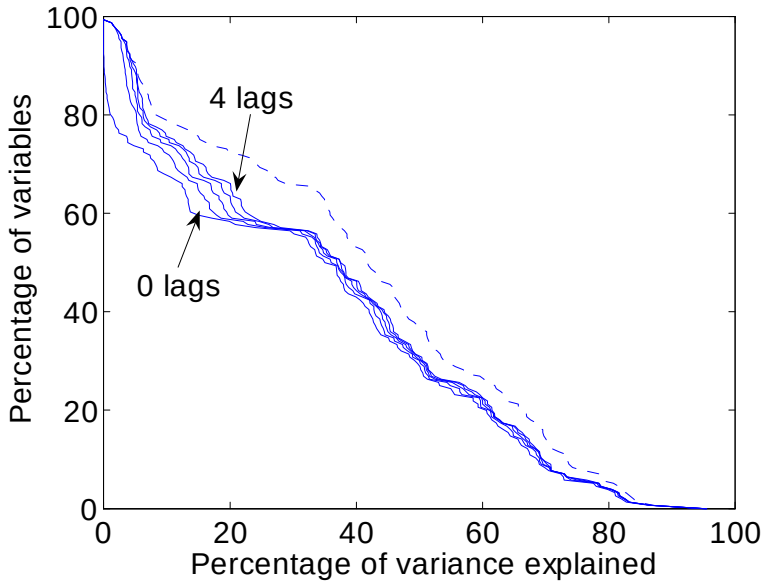

Figure 5: Pervasiveness of the effect of few lags only of the leisure preference shock on the observables.

we plot the graphs of the finite-lag pervasiveness functions $\tilde{y}_{lp}(z)$ defined as the percentage of the components $Y_{it}^{(s)}$ of $Y_t^{(s)}$ for which the variance of the linear projection on the space spanned by $f_{lp,t}, ..., f_{lp,t-h}$ constitutes at least $z\%$ of the total variance of $Y_{it}^{(s)}$.

Note that the area under the graph (divided by 100^2) equals the average fraction of the variance of $Y_{it}^{(s)}$ explained by $\tilde{\Lambda}_{i,lp}(L) f_{lp,t}$ with the average taken over $i = 1, 2, ..., 156$. We see that, in terms of the explanatory power, not much is lost by projecting the observables on the contemporaneous only ($h = 0$) leisure preference shock. To see this, compare the dashed and solid lines. As h increases, small improvements to the explanatory power take place. The largest improvement corresponds to the transition from $h = 0$ to $h = 1$.

Most of the changes to the graphs as h rises happen for $y > 60$. The reason is that this section of ordinates happens to correspond to sectoral inflations (section with $y > 80$) and sectoral wages ($80 > y > 60$). The contemporaneous leisure preference shock has essentially zero explanatory power for all the inflation indexes and very little explanatory power for sectoral wages. However, the one-period lagged leisure preference shock captures almost all the effect (still very small) of the entire history of the leisure preference shocks on inflation (the one-lag graph almost coincides with the dashed line when $y > 80$). It also considerably helps to explain the level of wages.[7]

[7] It is likely that one lag of the leisure preference shock would have almost as much of the explanatory power as the entire history of the leisure preference shock for the wage inflations (as opposed to the wage levels). We, however,

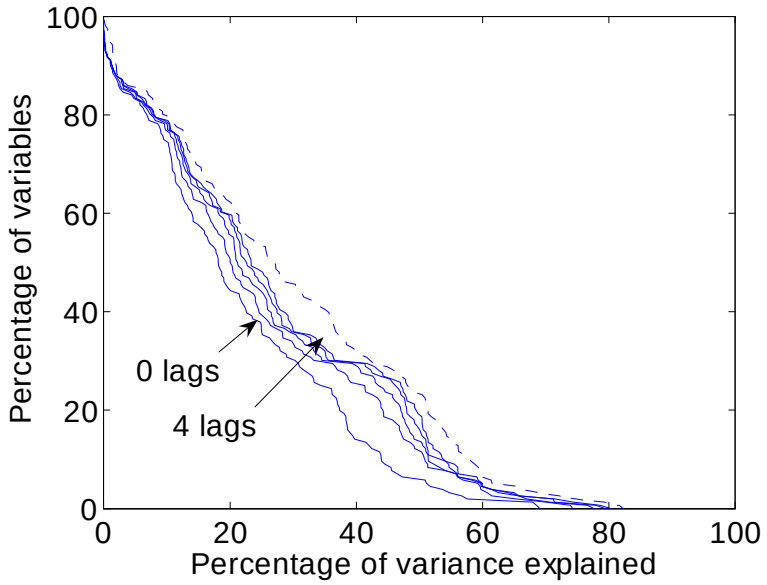

Figure 6: Pervasivenss of the effect of few lags only of the monetary policy shock on the observables.

Figures 6 and 7 are the equivalents of Figure 5 for the monetary policy and the money demand shocks, respectively. The improvement in the explanatory power when h increases from 0 to 1 is more substantial than in the case of the leisure preference shock. However, further increases in h do not increase the explanatory power of the monetary policy factor much and do not increase the explanatory power of the money demand factor at all. Overall, we conclude that replacing the entries of $\Lambda^{(s)}(L)$ with finite lag polynomials so that the entries corresponding to the leisure preference shock become scalars and the entries corresponding to the monetary policy and money demand shock become polynomials of degree one, would not do much harm to the explanatory power of $\Lambda^{(s)}(L)f_t$.

From the empirical perspective, our results provide support for the strategy of using a parsimonious number of lags to summarize the information about the observable variables that is contained in $\Lambda(L)f_t$ for the purpose of factor extraction and forecasting. However, the "correct" number of lags is an open question that requires the development of appropriate information criteria.

have included the levels rather than differences of wages in our dataset. As a result, there is a large discrepancy between the solid lines and the dashed line in the $80 > y > 60$ range of Figure 5.

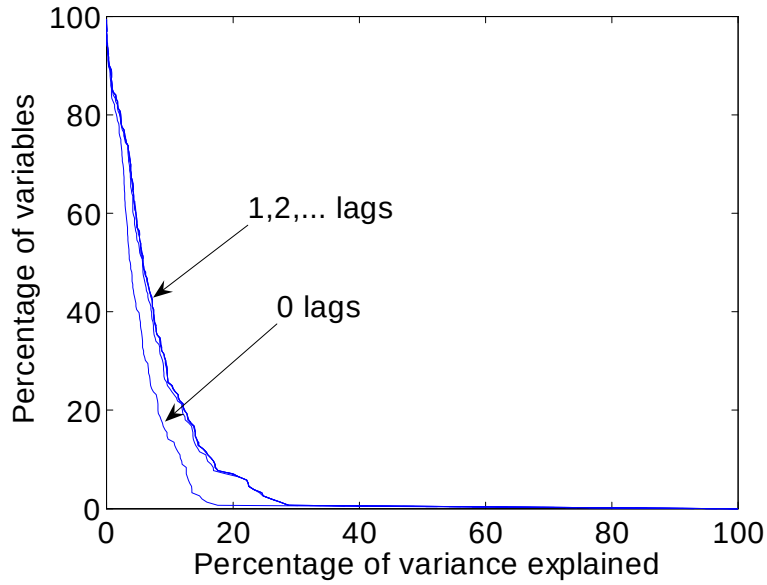

Figure 7: Pervasivenss of the effect of few lags only of the money demand shock on the observables.

4.1 The content of the principal components

Since, as we saw above, only a very few lags of f_t substantially add to factors' explanatory power, we can approximate the dynamic factor decomposition (4) by the following static factor decomposition. Let mp_t, md_t and lp_t denote the monetary policy, money demand and leisure preference components of f_t. We define a vector of static factors F_t as

$$F_t = (mp_t, mp_{t-1}, md_t, md_{t-1}, lp_t)'$$

and write:

$$Y_t^{(s)} = \Psi F_t + \eta_t,$$

where (Ψ_{i1}, Ψ_{i2}) is the vector of the coefficients of the linear projection of $Y_{it}^{(s)}$ on mp_t and mp_{t-1}; (Ψ_{i3}, Ψ_{i4}) is the vector of the coefficient of the linear projection of $Y_{it}^{(s)}$ on md_t and md_{t-1}; and Ψ_{i5} is the coefficient of the linear projection of $Y_{it}^{(s)}$ on lp_t. We do not include lp_{t-1} into F_t because, as can be seen from Figure 5, the lag of the leisure preference shock has very little explanatory power in our model. The lags of the monetary policy and money demand shocks have larger explanatory power and we include them into F_t.

In practice, F_t is not observed, and the space spanned by its components is estimated by the space spanned by the first five principal components of $Y_t^{(s)}$. Below, we study the difference between

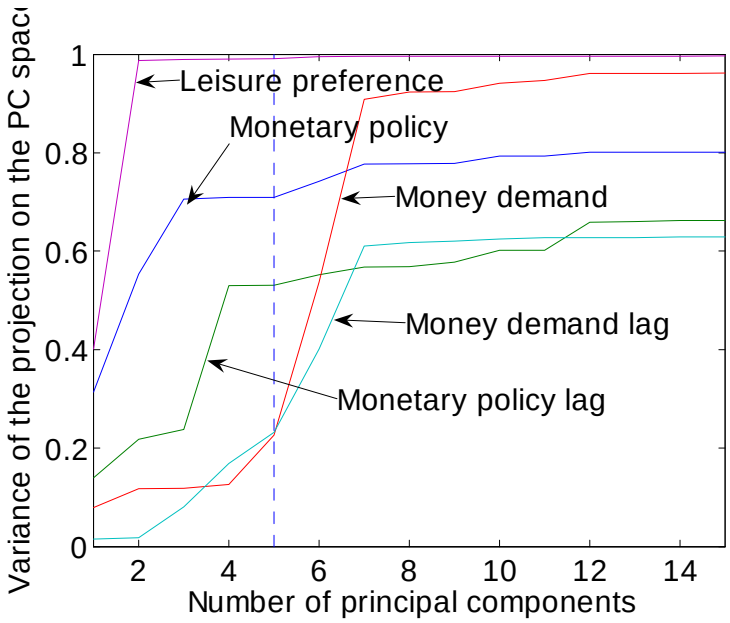

Figure 8: Portion of the variance of different components of the vector of static factors F_t explained by the projection on the spaces spanned by the first several (static, population) principal components.

the two spaces. We focus on the population principal components to see what the theoretical limit to the principal-component-based estimation of F_t is.

We would expect the principal components work well if all five non-zero eigenvalues of $\Psi E\left(F_t F_t'\right)\Psi'$ are substantially above the largest eigenvalue of $E\eta_t\eta_t'$. Unfortunately, this is not the case. Our computations show that the eigenvalues of $\Psi E\left(F_t F_t'\right)\Psi'$ equal 64.54, 30.49, 5.45, 1.46, and 0.003. But the largest eigenvalue of $E\eta_t\eta_t'$ equals 21.68, which is larger than three out of the five eigenvalues of $\Psi E\left(F_t F_t'\right)\Psi'$.

Then, what do the first five population principal components of $Y_t^{(s)}$ correspond to? Figure 8 reports the variances of the projections of the components of F_t (normalized to have unit variance each) on the space spanned by the first r population principal components of $Y_t^{(s)}$ as functions of r. Had the space of the first five principal components spanned the same space as the components of F_t, the variances of the projections would each be equal to one at $r = 5$.

From the figure, we see that the spaces spanned by the components of F_t and by the first five principal components are substantially different. Only the leisure preference shock component of F_t is accurately "recovered" by the principal components. The money demand shock content of the space of the principal components is particularly insignificant. However, the money demand shock

starts to be non-trivially present in the principal components' space when r becomes larger than six.

A possibility remains that the first five principal components span the same space as five linear filters of f_t, which are different from the $mp_t, mp_{t-1}, md_t, md_{t-1}$ and lp_t that we (somewhat subjectively) chose above to represent static factors. That this possibility does not realize can be seen from the following calculation. We compute the proportions of the variances of the first five principal components explained by their projections on the space spanned by the entire history of f_t. The computed proportions turn out to be: 96.6%, 98.3%, 80.8%, 76.9%, and 24.6% for the first, second, third, fourth, and fifth principal components, respectively.[8] Therefore, the fifth principal component has very little to do with the history of the economy-wide shocks.

4.2 The content of the generalized principal components

In this section, we would like to know whether the space of the generalized principal components proposed by Forni et al. (2005) better match the space of F_t than the space of the static principal components. Recall that the (population) generalized principal components are defined as follows. Let $S_Y(\omega)$ be the spectral density matrix of the standardized data and let $\lambda_j(\omega)$ and $p_j(\omega)$ be its j-th largest eigenvalue and the corresponding unit-length row eigenvector, respectively, so that $p_j(\omega) S_Y(\omega) = \lambda_j(\omega) p_j(\omega)$. Consider the sums

$$\begin{aligned} S^{\chi}(\omega) &\equiv \sum_{j=1}^{q} \lambda_j(\omega) \overline{p_j(\omega)}' p_j(\omega) \\ S^{\xi}(\omega) &\equiv \sum_{j=q+1}^{n} \lambda_j(\omega) \overline{p_j(\omega)}' p_j(\omega) \end{aligned}$$

where q is the number of dynamic factors, and compute

$$\begin{aligned} \Gamma_0^{\chi} &= \int_{-\pi}^{\pi} S^{\chi}(\omega)\, d\omega \\ \Gamma_0^{\xi} &= \int_{-\pi}^{\pi} S^{\xi}(\omega)\, d\omega. \end{aligned}$$

We set $q = 3$ because there are three economy-wide shocks in the model. Then, the j-th (population) generalized principal component of $Y_t^{(s)}$ relative to the couple $\left(\Gamma_0^{\chi}, \Gamma_0^{\xi}\right)$ is defined as $Z_j Y_t^{(s)}$, where Z_j are the solutions of the generalized eigenvalue equations

$$Z_j \Gamma_0^{\chi} = \nu_j Z_j \Gamma_0^{\xi}, \tag{5}$$

[8] Compare these to the proportions of the variances of the first five principal components explained by their projections on the space spanned by $mp_t, mp_{t-1}, md_t, md_{t-1}$ and lp_t only: 82.3%, 88.5%, 41.8%, 45.3%, and 10.5%.

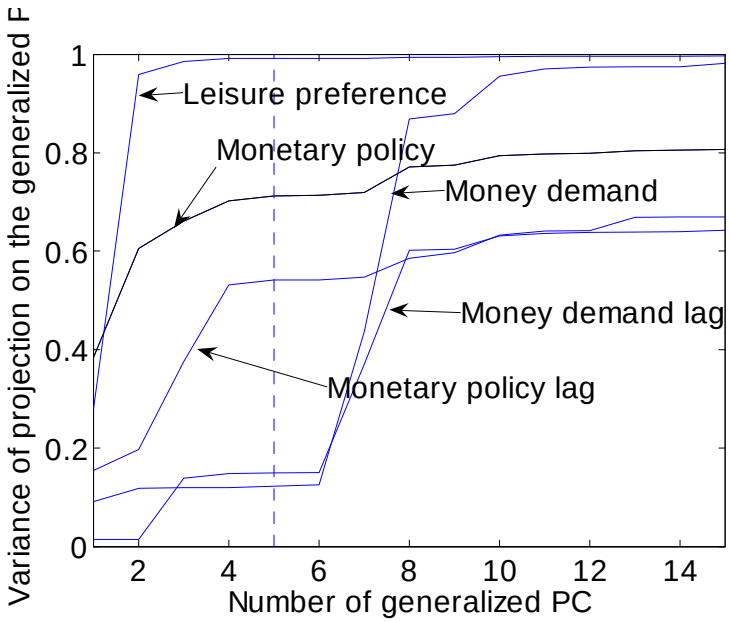

Figure 9: Portion of the variance of different components of the vector of static factors F_t explained by the projection on the spaces spanned by the first several (population) generalized principal components.

for $j = 1, 2, ..., n$, with the normalization constraints $Z_j \Gamma_0^\xi Z_j' = 1$ and $Z_i \Gamma_0^\xi Z_j' = 0$ for $i \neq j$, and with $\nu_1 \geq \nu_2 \geq ... \geq \nu_n$.

For our data generating process, the problem (5) is ill-posed in the sense that very small changes in Γ_0^ξ may result in large changes in the solutions. This is so because the matrix Γ_0^ξ turns out to have a large number of eigenvalues numerically close to zero. One way to proceed, which Forni et al. (2005) choose to follow, is to replace Γ_0^ξ by the matrix with the same diagonal, but with zero off-diagonal elements. We do such a replacement in what follows.

Figure 9 is the equivalent of Figure 8 for the case of the generalized principal components. The two figures are very similar. The proportions of the variances of the first five generalized principal components explained by their projections on the space spanned by the entire history of f_t are equal to 89.5%, 93.2%, 55.9%, 55.9% and 5.5%. These figures suggest somewhat smaller macroeconomic content of the generalized principal components relative to the static principal components, for which the analogous figures reported above are: 96.6%, 98.3%, 80.8%, 76.9%, and 24.6%.[9]

We conclude this section by summarizing its main finding: for our data generating process,

[9] The proportions of the variances of the first five generalized principal components explained by their projections on the space spanned by $mp_t, mp_{t-1}, md_t, md_{t-1}$ and lp_t only are: 73.3%, 91.9%, 42.1%, 38.4%, and 4.5%. Compare these to the analagous numbers 82.3%, 88.5%, 41.8%, 45.3%, and 10.5% for the static principal components.

the information about the space of the macroeconomic shocks and their most important lags is scattered through a relatively large number of principal components. Knowing the first five principal components is not sufficient to accurately recover the five shocks and lags. Knowing more principal components helps such a recovery. For example, in our exercises, the portion of the variance of the money demand shock and its lag explained by their projections on the space spanned by the first r principal components increase from below 20% for $r = 5$ to more than 60% for $r = 7$ or (in case of the generalized PC) $r = 8$. The principal component space of the same dimension as the number of shocks and their important lags may poorly approximate the "macroeconomic space".

5. Number of factors

The true number of factors may be interpreted in many different ways in the DSGE model of Bouakez et al. (2009). In this sense, the model replicates the same ambiguity that we find in actual applied research. Depending on the goal of the analysis, we might want to estimate different number of factors. Sometimes, the relevant goal of the analysis is to determine the number of basic macroeconomic shocks influencing the dynamics of a large number of macroeconomic indicators. In this section, we therefore ask the following question. What is the relationship between the number of dynamic factors estimated from the data simulated from the model and the number of the economy-wide shocks in this model, which is three?

We use the Bai-Ng (2007) and Hallin-Liska (2007) criteria to estimate the number of factors in 1000 different simulations of our data with $n = 156$ and $T = 120$. Before applying the criteria, we demean and standardize the simulated data. First, we apply the Hallin-Liska (2007) method. The choice of the tuning parameters of the Hallin-Liska method is as follows. In Hallin and Liska's (2007) notation, we use the information criterion $IC^T_{2;n}$ with penalty $p_1(n, T)$, set the truncation parameter M_T at $\left[0.7\sqrt{T}\right]$ and consider the subsample sizes $(n_j, T_j) = (n - 10j, T - 10j)$ with $j = 0, 1, 2, 3$ so that the number of the subsamples is $J = 4$. The cross-sectional units excluded from the subsamples are determined randomly. We chose the penalty multiplier c on a grid $0.01 : 3$ wiht increment 0.01 using Hallin and Liska's second "stability interval" procedure. We applied the Hallin-Liska method to 1000 different simulations of our data, choosing the subsamples randomly each time and setting the maximum number of dynamic factors at 8. Out of 1000 times, the method finds one dynamic factor 4 times, two dynamic factors 100 times, three dynamic factors 274 times, four dynamic factors 259 times, five dynamic factors 180 times, six dynamic factors 109 times, and seven dynamic factors 74 times.

Next, we use the Bai-Ng (2007) method to determine the number of dynamic factors. We set the maximum number of static factors at 10, and, in the notation of Bai and Ng (2007), use either

$\hat{D}_{1,k}$ or $\hat{D}_{2,k}$ statistic for either the residuals or the standardized residuals of a VAR(4) fitted to the estimated factors, and consider $\delta = 0.1$ and $m = 2, 1$ and 0.5. Table 1 summarizes our findings for different choices of the tuning parameters.

Table 1: Number of times different estimates of the number of factors appeared in 1000 simulations

Residuals	Not standardized						Standardized					
Statistic used	$\hat{D}_{1,k}$			$\hat{D}_{2,k}$			$\hat{D}_{1,k}$			$\hat{D}_{2,k}$		
m	2	1	0.5	2	1	0.5	2	1	0.5	2	1	0.5
Estimate	Number of times appeared in 1000 simulations											
1	1000	1000	754	1000	1000	437	1					
2			246			555	368			3		
3						8	595	18		342	1	
4							36	330	1	614	49	
5								555	85	41	542	10
6								97	495		392	292
7									385		16	620
8									34			78

Bai and Ng estimates based on the standardized residuals are less conclusive than their estimates based on raw residuals, which suggest that there is one, or perhaps, two dynamic factors in the data. When the standardized residuals are used in the Bai and Ng procedure, the estimated number of dynamic factors depends very much on the parameter m, which regulates the scale of the threshold below which the eigenvalues of the sample correlation matrix of residuals are interpreted as small enough to conclude that the corresponding eigenvalues of the population correlation matrix of the errors of the VAR(4) equal zero. The dependence of the conclusions on the choice of the threshold was the motivation for Hallin and Liska to design their second stability interval procedure. In our Monte Carlo experiment, the Hallin-Liska estimates of the number of factors vary, the most frequent estimates being three, four and five dynamic factors.

Finally, we test different hypotheses about the number of dynamic factors using Onatski's (2009) test. Figure 10 reports the cumulative empirical distributions of the p-values computed in 1,000 Monte Carlo experiments for the test of 0 factors vs. 1 factor, 1 factor vs. 2 factors, 2 factors vs. 3 factors, and so forth. The steeply increasing cumulative distribution function (c.d.f.) corresponds to the null of zero factors, which indicates that the probability of rejection of this null is much larger than the size of the test. The next steepest c.d.f. is for the null of one factor. Note that we are still rejecting the null hypothesis with probability much larger than the size. The third steepest c.d.f. is not steep. So if we were to perform the test in a single "typical" monte Carlo simulation of the data, we would likely accepted either the hypothesis of one dynamic factor, or the hypothesis

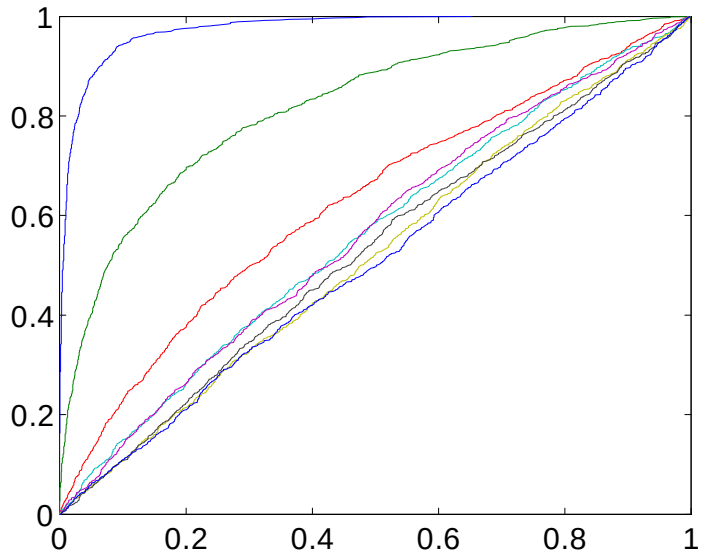

Figure 10: The Monte Carlo cumulative empirical distributions of the p-values for Onatski's (2009) test of 0 factors vs. 1 factor, 1 factor vs. 2 factors, 2 factors vs. 3 factors, etc.

of two dynamic factors.

All the above methods of the determination of the number of factors are based on the idea that the number of exploding eigenvalues of the spectral density matrix of the data as $n \to \infty$ equals the true number of factors. However, as Figure 3 shows, only two of the "economy-wide eigenvalues" are larger than the largest "idiosyncratic eigenvalue". Hence, whenever the above criteria estimate more than two factors, the reason for such an estimate is not that the criterion is sensitive enough to detect the third macroeconomic shock from the noisy signal. The reason is simply that the criterion classifies some sector-specific sources of variation as influential enough to call them factors.

As was mentioned above, in practice, the dynamic factor models are often represented in the static form and then estimated by principal components. Before the estimation, the number of static factors should be determined. As in the case of the dynamic factors, the number of static factors can be interpreted in many different ways, and different loss functions imply different optimal estimates of this number. If the goal is to determine the number of macroeconomic shocks and their lags which suffice to accurately describe the systematic components of the data, then, perhaps, the desired estimate is five as discussed above.

A basic informal method for the determination of the number of factors was proposed by Cattel (see Cattel, 1966). It uses the visual analysis of the scree plot, which is the line that connects the

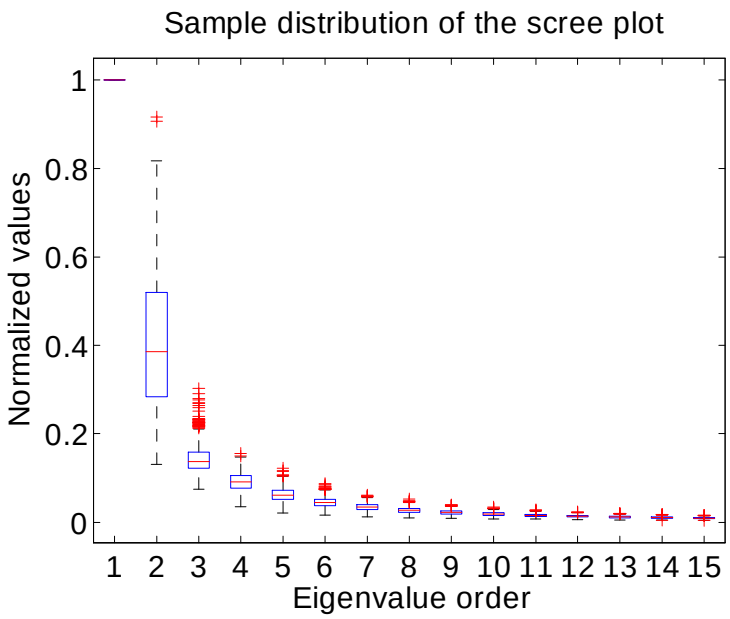

Figure 11: Monte Carlo distribution of the scree plot for the data simulated from the BCR model.

decreasing eigenvalues of the sample covariance matrix of the data plotted against their respective order numbers. In practice, it often happens that the scree plot shows a sharp break where the true number of factors ends and "debris" corresponding to the idiosyncratic influences appears.

Figure 11 shows the 15 largest eigenvalues (normalized so that the largest eigenvalue is 1) of the sample correlation for our simulated dataset. The boxplots represent the sample distribution of the 15 largest eigenvalues (based on 1000 Monte Carlo simulations). By looking at the scree, a researcher would sometimes think that there is only one static factor in the data and sometimes that there are two or three such factors.

Next, we apply Bai and Ng's (2002) criteria PC_{p1}, PC_{p2}, PC_{p3}, IC_{p1}, IC_{p2}, IC_{p3}, and BIC_3, and Onatski's (2005) ED criterion to determine the number of static factors in each of the 1,000 Monte Carlo simulations of our data generating process (DGP).[10] We consider three choices of the maximum number of static factors: $r_{\max} = 5$, $r_{\max} = 10$ and $r_{\max} = 15$. For all the three choices, the criteria PC_{p1}, PC_{p2}, PC_{p3}, IC_{p1}, IC_{p2} and IC_{p3} estimate the number of static factors equal to $r_{\max}$. Criteria BIC_3 and ED produce estimates which are smaller than $r_{\max}$ in most of the considered cases.

The sample distributions of the estimates corresponding to BIC_3 and ED (in the Monte Carlo sample of 1,000 simulations) are shown in Figure 12. The number of factors estimated by BIC_3,

[10] Again, we demean and standardize the simulated data prior to applying Bai and Ng's criteria and ED.

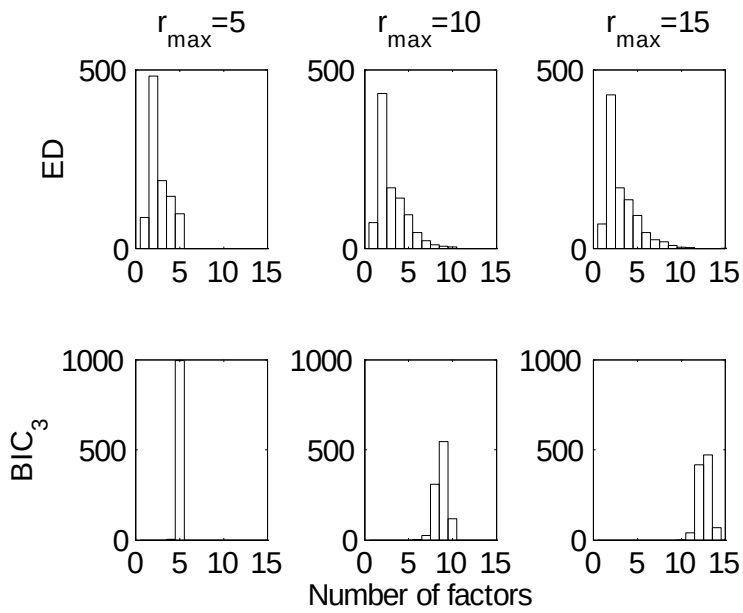

Figure 12: The Monte Carlo distributions of the BIC_3 and ED estimates of the number of static factors.

although smaller than $r_{\max}$ in most of the cases, still very much depends on $r_{\max}$. In contrast, the results of ED are rather insensitive to $r_{\max}$. In most of the Monte Carlo experiments, the number of static factors estimated by ED is either 2 or 3.

Our findings so far can be summarized as follows. For our DGP, the economy-wide shocks do have a pervasive effect on a large number of generated variables. However, the effects of these shocks are not heterogeneous enough to identify the space of the macroeconomic fluctuations with the space of the few linear filters of the data explaining most of its variance. Although most of the common dynamics of the data can be explained by the current economy-wide shocks and a very few of their lags, the space spanned by these shocks and lags is substantially different from the space of the population principal components of the same dimensionality. The first few of the principal components depend almost entirely on the macroeconomic shocks and their lags. However, the more distant principal component have a large sector-specific content. The information about the space of the economy-wide shocks and their lags is spread through a relatively large number of the population principal components.

6. Sample *vs.* population principal components

In practice, factors are estimated by sample principal components. Hence, even if the spaces of factors and population principal components coincided, there would be a discrepancy between the true and estimated factor space because sample and population principal components differ. In this section, we ask how large such a discrepancy would be for our data generating process. To answer this question, we redefine the true static factors of our standardized data as their population principal components. That is $F_{jt} = (1/\sqrt{\lambda_{0j}})v_{0j}'Y_t^{(s)}$, where λ_{0j} is the j-th largest eigenvalue and v_{0j} is the corresponding eigenvector of the population covariance matrix of $Y_t^{(s)}$. As before, we simulate 1,000 datasets $Y_1, ..., Y_{120}$. For each simulation, we compute matrix $\frac{1}{120}\sum_{t=1}^{120} Y_t^{(s)}Y_t^{(s)\prime}$, its eigenvalues λ_j and the corresponding eigenvectors v_j.[11] Then the sample principal components are: $\hat{F}_{jt} = (1/\sqrt{\lambda_j})v_j'Y_t^{(s)}$. They estimate the "true" static factors F_{jt}.

6.1 Regressions of $\hat{F}$ on F

First, for each of $\hat{F}_{1t}, ..., \hat{F}_{6t}$, we compute the average R^2 in the regression of $\hat{F}_{jt}$ on the constant and $F_{1t}, ..., F_{6t}$. For $j = 1, ..., 6$, the corresponding numbers are: 1, 1, 0.98, 0.91, 0.73, and 0.57. We clearly see that the more distant static factors are more poorly estimated. When we regress $\hat{F}_{jt}$ on F_{jt} and a constant only, we get the following average R^2 for $j = 1, ..., 6$: 0.89, 0.65, 0.67, 0.49, 0.22, and 0.16. Note that the higher average R^2 in the regressions of $\hat{F}_{jt}$ on $F_{1t}, ..., F_{6t}$ relative to the regressions of $\hat{F}_{jt}$ on F_{jt} is not a consequence of the fact that the "true" static factors are identified only up to a non-singular transformation. In our case, they are identified up to a sign because we define them as the population principal components. This phenomenon deserves further serious exploration, which we leave for future research.

The substantial decrease in the R^2 with j in the individual regressions of $\hat{F}_{jt}$ on F_{jt} is consistent with Onatski's (2005) finding that the coefficient in the regression of a principal component estimate of a weak factor on the factor itself is substantially biased towards zero. The weaker the factor, the larger the bias and the smaller the R^2.

6.2 Accuracy of the asymptotic approximation

Next, we consider the accuracy of Jushan Bai's (2003) asymptotic approximation to the finite sample distribution of $\hat{F}_t = \left(\hat{F}_{1t}, ..., \hat{F}_{kt}\right)'$. Let X be an n by T matrix of our simulated stan-

[11] In practice, the standardized data would be obtained by using estimated standard deviations of the raw data series. Here, however, we abstract from this fact and use $Y_t^{(s)}$, which are standardized by the true standard deviations. This allows us to focus on the difference between the population and sample principal components caused only by the difference between the eigenstructures of $EY_t^{(s)}Y_t^{(s)\prime}$ and $\frac{1}{120}\sum_{t=1}^{120} Y_t^{(s)}Y_t^{(s)\prime}$.

dardized data $\left[Y_1^{(s)}, ..., Y_T^{(s)}\right]$. Let $F_t = (F_{1t}, ..., F_{kt})'$ be the true factors defined as the population principal components, let $F = [F_1, ..., F_T]'$, and let Λ^0 be the true factor loadings, that is $\Lambda^0 = [\sqrt{\lambda_{01}} v_{01}, ..., \sqrt{\lambda_{0k}} v_{0k}]$. Bai's Theorem 1 says that, as long as certain asymptotic regularity conditions are satisfied[12] and $\sqrt{n}/T \to 0$, for each t, we have:

$$\sqrt{n}\left(\hat{F}_t - H'_{n,T} F_t\right) \xrightarrow{d} N\left(0, \Pi_t\right), \tag{6}$$

where

$$H_{n,T} \equiv \frac{\Lambda^{0\prime}\Lambda^0}{n} \frac{F'\hat{F}}{T} V_{nT}^{-1},$$

$V_{nT} \equiv diag\left(\hat{\lambda}_1, ..., \hat{\lambda}_k\right)$ is the diagonal matrix of the first k of the largest eigenvalues of $XX'/(nT)$, $\Pi_t = V^{-1} Q \Gamma_t Q' V^{-1}$, V is the diagonal matrix of eigenvalues of $\Sigma_\Lambda^{1/2} \Sigma_F \Sigma_\Lambda^{1/2}$, Σ_F being the probability limit of $F'F/T$ and Σ_Λ being the probability limit of $\Lambda^{0\prime}\Lambda^0/n$, $Q = V^{1/2} \Psi' \Sigma_\Lambda^{-1/2}$ with Ψ having columns equal to the normalized eigenvectors of $\Sigma_\Lambda^{1/2} \Sigma_F \Sigma_\Lambda^{1/2}$, $\Gamma_t = \lim_{n\to\infty}(1/n) \sum_{i=1}^n \sum_{j=1}^n \Lambda_i^0 \Lambda_j^{0\prime} E e_{it}^{(s)} e_{jt}^{(s)}$ and $e_{it}^{(s)}$ are the idiosyncratic components in the factor model.

In fact, the above asymptotics used in our setting implies that

$$\sqrt{n}\left(\hat{F}_t - H'_{n,T} F_t\right) \xrightarrow{p} 0. \tag{7}$$

This is because $\Pi_t = 0$. Indeed, let us denote the matrix $[v_{01}, v_{02}, ..., v_{0k}]$ as V_{0k} and the matrix $diag\left(\lambda_{01}, \lambda_{02}, ..., \lambda_{0k}\right)$ as S_{0k}. Then, $\Lambda^0 = V_{0k} S_{0k}^{1/2}$ and $F_t = S_{0k}^{-1/2} V_{0k}' Y_t^{(s)}$. Since n is fixed (but large) in our experiment, we cannot meaningfully take the limit of $(1/n) \sum_{i=1}^n \sum_{j=1}^n \Lambda_i^0 \Lambda_j^{0\prime} E e_{it}^{(s)} e_{jt}^{(s)}$ as $n \to \infty$. The best we can do is to set $\Gamma_t = (1/n) \sum_{i=1}^n \sum_{j=1}^n \Lambda_i^0 \Lambda_j^{0\prime} E e_{it}^{(s)} e_{jt}^{(s)}$. But $E e_t^{(s)} e_t^{(s)\prime} = E\left(Y_t^{(s)} Y_t^{(s)\prime}\right) - V_{0k} S_{0k} V_{0k}'$. Therefore,

$$\begin{aligned}
\Gamma_t &= (1/n)\Lambda^{0\prime}\left[E\left(Y_t^{(s)} Y_t^{(s)\prime}\right) - V_{0k} S_{0k} V_{0k}'\right]\Lambda^0 \\
&= (1/n) S_{0k}^{1/2} V_{0k}' \left[E\left(Y_t^{(s)} Y_t^{(s)\prime}\right) - V_{0k} S_{0k} V_{0k}'\right] V_{0k} S_{0k}^{1/2} \\
&= (1/n) S_{0k}^{1/2} \left[S_{0k}^{1/2} E\left(F_t F_t'\right) S_{0k}^{1/2} - S_{0k}\right] S_{0k}^{1/2} \\
&= (1/n) S_{0k} \left[E\left(F_t F_t'\right) - I_k\right] S_{0k} = 0.
\end{aligned}$$

We can check how accurate the asymptotic approximation (7) is in our case. Following the above Monte Carlo structure, we simulated 1,000 datasets and computed $\sqrt{n}\left(\hat{F}_t - H'_{n,T} F_t\right)$ for each Monte Carlo replication. Figure 13 shows the median, and the 25 and 75 percentiles of the empirical distribution of the 1,000 Monte Carlo replications of $\sqrt{n}\left(\hat{F}_t - H'_{n,T} F_t\right)$ when only one factor is estimated. We see that the approximation (7) works very poorly. Not only the interquartile

[12] These conditions are asymptotic and, strictly speaking, cannot be checked in any finite sample such as ours.

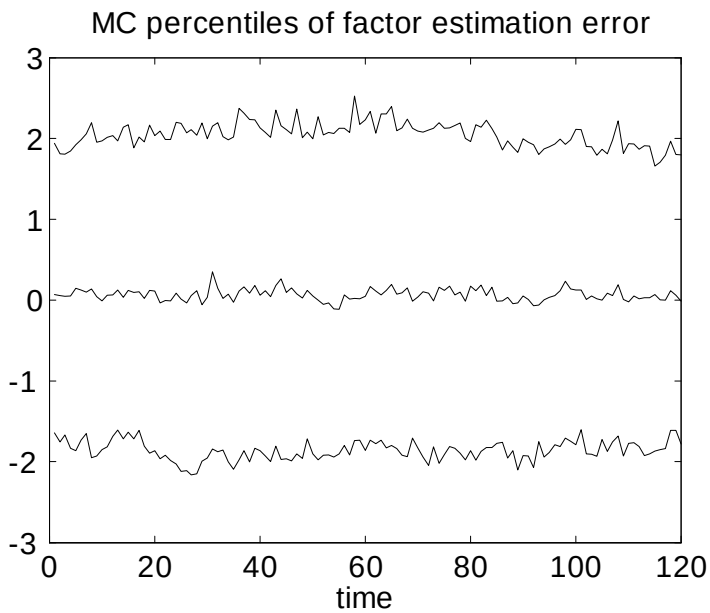

Figure 13: The interquartile range and the median of the empirical distribution of the 1,000 Monte Carlo replications of $\sqrt{n}\left(\hat{F}_t - H'_{n,T}F_t\right)$.

range of the empirical distribution of the Monte Carlo replications is wide, the distribution also has very fat tails in the sense that the moments of this distribution are large. For example, the standard deviation of $\sqrt{n}\left(\hat{F}_t - H'_{n,T}F_t\right)$ is more than 4 for most of $t = 1, 2, ..., 120$.

Of course, if a researcher only had data simulated from our model but did not know the model itself, she would not know that $\Pi_t = 0$ and would estimate it. In his Monte Carlo experiments, Bai (2004) estimates Π_t by $\hat{\Pi}_t = V_{nT}^{-1}(1/n)\sum_{i=1}^{n}\hat{e}_{it}^2\hat{\Lambda}_i\hat{\Lambda}_i'V_{nT}^{-1}$, where $\hat{e}_{it} = X_{it} - \hat{\Lambda}_i\hat{F}_t'$ and $\hat{\Lambda}' = \hat{F}'X/T$. To assess the quality of the asymptotic approximation (6), Bai (2004) computes the standardized estimate

$$g_t = \hat{\Pi}_t^{-1/2}\sqrt{n}\left(\hat{F}_t - H'_{n,T}F_t\right) \tag{8}$$

and compares the empirical distribution of g_t (for a fixed t) with the standard normal distribution.

We use Bai's strategy below. For the moment, we consider the case of only one factor. One problem with such an exercise would be a double (imperfect) cancellation of errors. The quantity $\sqrt{n}\left(\hat{F}_t - H'_{n,T}F_t\right)$ would not be close to zero as we saw from Figure 13, but $\hat{\Pi}_t$ will also be far from zero because $(1/n)\sum_{i=1}^{n}\hat{\Lambda}_i\hat{\Lambda}_i'\hat{e}_{it}^2$ would not be close to $(1/n)\sum_{i=1}^{n}\sum_{j=1}^{n}\Lambda_i^0\Lambda_j^{0\prime}Ee_{it}^{(s)}e_{jt}^{(s)}$ since the cross-sectional serial correlation is ignored and since T is not much larger than n.

When we plot the median and 25% and 75% percentiles of the Monte Carlo distribution of g_t as functions of $t = 1, ..., T = 120$, the resulting graph turns out to be similar to that on Figure 13. To

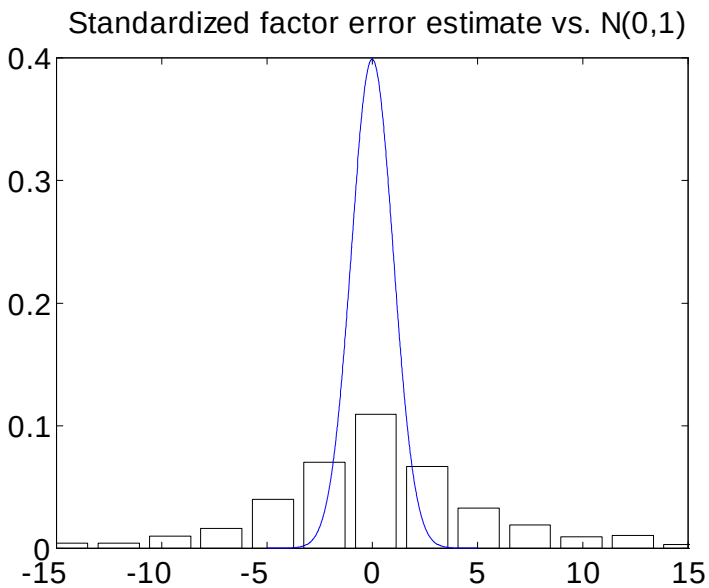

Figure 14: The Monte Carlo distribution of g_{60} relative to the distribution of the standard normal random variable.

save space, we do not report this graph. Instead, we show the histogram for the 1,000 replications of f_t when $t = T/2 = 60$ on Figure 14. The histogram is scaled so that the sum of the areas of all the bars equals 1 and interimposed with the standard normal density.[13]

When several factors are estimated, the situation remains bad. The way in which the asymptotic approximation is not accurate varies with the explanatory power of the factor. Figure 15 shows the histograms for the Monte Carlo distributions of the components of g_t when four factors are estimated. We see that the standard normal asymptotic approximation to the empirical distribution of g_{1t} is relatively accurate, but the empirical distributions of g_{2t}, g_{3t} and g_{4t} have progressively fatter tails relative to the standard normal approximation. We link such a deterioration of the quality of Bai's asymptotic approximation to the decreasing explanatory power of more distant factors.[14]

We conclude that the asymptotic theory of the principal components estimates available to date does not provide a good approximation to the distribution of the estimates for our data

[13] The range of the horizontal axis is truncated at [-15,15] to improve visibility.

[14] Onatski (2005) describes a different asymptotic approximation designed specifically for the case when the factors are weak, as seems to be the case in our present exercise. However, Onatski's formulas are derived for the case of neither cross-sectional nor serial correlation in the idiosyncratic terms and cannot be used here. In our opinion, derivation of the alternative weak factor asymptotics for the case of correlated idiosyncratic terms is an important task. We leave it for future research.

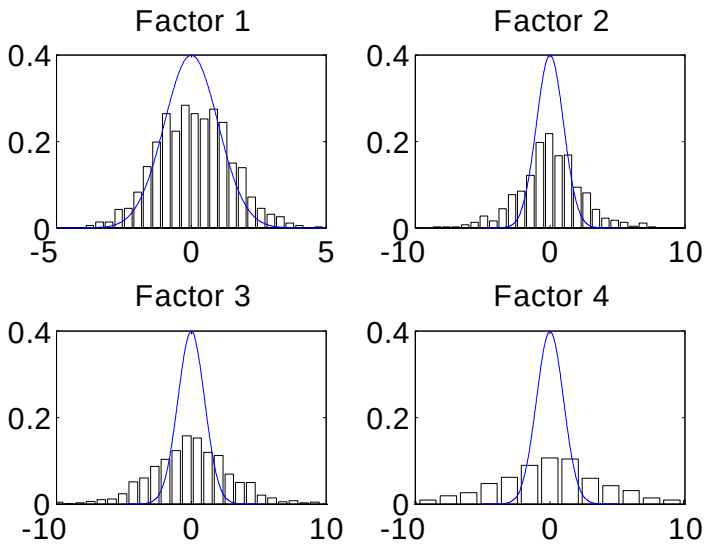

Figure 15: Monte Carlo distributions of $g_{1,60}$, $g_{2,60}$, $g_{3,60}$, and $g_{4,60}$ relative to the theoretically derived standard normal distribution.

generating process. To the extent that the data generating process we consider is similar to actual macroeconomic datasets, this finding calls for a new asymptotic theory, perhaps, based on the assumptions which better correspond to the finite sample situation.

7. Diffusion Index Forecasting

Large factor models are often used as statistical devices which motivate and explain diffusion index forecasting. Stock and Watson (2002) propose the following factor model framework for forecasting variable y_{t+1} using a high-dimensional vector Y_t of predictor variables:

$$y_{t+1} = \beta(L) f_t + \gamma(L) y_t + \varepsilon_{t+1} \tag{9}$$

$$Y_{it} = \lambda_i(L) f_t + e_{it}, \tag{10}$$

where $\beta(L), \gamma(L)$ and $\lambda_i(L)$ are lag polynomials in non-negative powers of L, f_t are a few economy-wide shocks influencing a wide range of variables Y_{it}, and $E(\varepsilon_{t+1}|f_t, y_t, Y_t, f_{t-1}, y_{t-1}, Y_{t-1}, ...) = 0$.

It turns out that equations (9)-(10) fit in our theoretical model quite well. For example, if we set y_t equal to output growth or to aggregate inflation, and if we set f_t equal to the three-dimensional vector of economy-wide shocks: the monetary policy, the money demand and the leisure preference shocks, then the linear forecasts of y_{t+1} based on the history of y_t and f_t are nearly optimal so that, although the assumption $E(\varepsilon_{t+1}|f_t, y_t, Y_t, f_{t-1}, y_{t-1}, Y_{t-1}, ...) = 0$ is violated because

$E(y_{t+1}|f_t, y_t, f_{t-1}, y_{t-1}, ...) \neq E(y_{t+1}|f_t, y_t, Y_t, f_{t-1}, y_{t-1}, Y_{t-1}, ...)$, the degree of the violation is very small. Precisely, using the state space representation (1) for our model, we numerically compare the variance of the optimal forecast error $\varepsilon^o_{t+1} = y_{t+1} - E(y_{t+1}|f_t, y_t, Y_t, f_{t-1}, y_{t-1}, Y_{t-1}, ...)$ with the variance of the forecast error $\varepsilon_{t+1} = y_{t+1} - E(y_{t+1}|f_t, y_t, f_{t-1}, y_{t-1}, ...)$. For output growth we find that

$$\frac{Var(y_{t+1}) - Var(\varepsilon_{t+1})}{Var(y_{t+1}) - Var(\varepsilon^o_{t+1})} = 0.991, \tag{11}$$

which means that the history of y_t and f_t alone captures 99.1% of all forecastable variance in y_{t+1}. For aggregate inflation, the corresponding figure is 98.9%.

The reason why we use the above measure of forecast accuracy as opposed to, say, the mean squared forecast error ratio $Var(\varepsilon_{t+1})/Var(\varepsilon^o_{t+1})$ is that in our model, output growth and aggregate inflation are not easily forecastable. For example, for output growth, the ratio $Var(\varepsilon^o_{t+1})/Var(y_{t+1})$ equals 0.8484 (for inflation, the similar number is 0.8616) so that only 15.2% of the variance of output growth (and 15.8% of the variance of aggregate inflation) is forecastable at one quarter horizon. Therefore, had we used $Var(\varepsilon_{t+1})/Var(\varepsilon^o_{t+1})$ as a measure of sub-optimal forecast accuracy, any reasonable forecast would produce a number in between 1 and $(0.8484)^{-1} = 1.179$ (the latter number would correspond to zero forecast) making the measure not particularly informative.

In our model, the factors represented by the three economy-wide shocks have large forecasting power relative to other predictors. For example, the forecast of output growth based on its own history and the history of aggregate inflation captures 79% of the forecastable variance, whereas the forecast based on the history of f_t alone captures 98.8% of the forecastable variance. For aggregate inflation, its forecast based on the own history and the history of output growth captures 75.5% of the forecastable variance, whereas the forecast based on the history of f_t alone captures 97.4% of the forecastable variance. Hence, the diffusion index forecasts substantially outperform small VAR forecasts in our model, at least theoretically. Further analysis of the predictive power of f_t shows that the forecasts based on the separate histories of the monetary policy shocks, money demand shocks and leisure preference shocks capture 47.8%, 46.6% and 4.4%, respectively, of the forecastable variance of output growth. For aggregate inflation, the similar numbers are 75.5%, 11.6% and 10.2%.

In practice, to implement (9)-(10), an important simplifying assumption that the lag polynomials $\beta(L)$, $\gamma(L)$ and $\lambda_i(L)$ are of finite order is usually made. Equation (10) is replaced by a static factor representation: $Y_{it} = \Lambda_i F_t + e_{it}$, and equation (9) is replaced by

$$y_{t+1} = \alpha_1 + \beta_1(L)F_t + \gamma_1(L)\, y_t + \varepsilon_{t+1}, \tag{12}$$

where $\beta_1(L)$ and $\gamma_1(L)$ are polynomials of finite order. Then, the static factors F_t are estimated

by the sample principal components and the coefficients of the above forecasting equation are estimated by ordinary least squares (OLS) of y_{t+1} on $\hat{F}_t, y_t$ and a few of their lags.

We have already seen that the space of principal component is substantially different from the space of macroeconomic static factors represented by $mp_t, mp_{t-1}, md_t, md_{t-1}$ and lp_t. There are substantial differences both at population and sample levels. At the population level, the macroeconomic space is substantially different from the space of the population principal components. At the sample level, the sample principal components imperfectly estimate the space of the population principal components (but, as is seen from Figure 15, the discrepancy is mostly due to the estimation of the less influential population principal components).

Although the difference between the macroeconomic and principal component spaces may be harmful for structural analysis, it may not matter for diffusion index forecasts. In this subsection, we explore this possibility in detail. We focus on the special case of (12):

$$y_{t+1} = \alpha + \beta' F_t + \varepsilon_{t+1}, \tag{13}$$

with no lags of y_{t+1} and no lags of F_t as predictor variables. We set[15]

$$F_t = (mp_t, mp_{t-1}, md_t, md_{t-1}, lp_t, lp_{t-1}). \tag{14}$$

Let us denote the vector of the first six population principal components of vector $Y_t^{(s)}$ as PC_t, and let us denote the vector of the first six sample principal components of $Y_t^{(s)}$, $t = 1, ..., T$, as $\widehat{PC}_t$. We estimate the coefficients α and β in (13) by the OLS regression of y_{t+1}, $t = 1, ..., T-1$ on a constant and $\widehat{PC}_t$, and use the estimates to form the forecast:

$$\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\beta}' \widehat{PC}_T.$$

Then, we decompose the forecast error into four different parts as follows:

$$\begin{aligned} y_{T+1} - \hat{y}_{T+1|T} \quad = \quad & \underbrace{y_{T+1} - E\left(y_{T+1}|Y_T, Y_{T-1}, ...\right)}_{\delta_1} + \\ & \underbrace{E\left(y_{T+1}|Y_T, Y_{T-1}, ...\right) - E^*\left(y_{T+1}|F_T\right)}_{\delta_2} + \\ & \underbrace{E^*\left(y_{T+1}|F_T\right) - E^*\left(y_{T+1}|PC_T\right)}_{\delta_3} + \\ & \underbrace{E^*\left(y_{T+1}|PC_T\right) - \hat{y}_{T+1|T}}_{\delta_4}. \end{aligned}$$

[15] We have included lp_{t-1} in the set of static factors because, as was mentioned in the discussion of Figure 5, it helps to explain the variance of inflation series.

The first part (δ_1) is the completely unforecastable part, which is the error of the best mean squared forecast of y_{T+1} given the entire history of the observables. The second part (δ_2) is the difference between the best forecast and the linear forecast based on the current value of the six static factors defined by (14). From the discussion above, we know that the linear forecast based on the entire history of f_t (and hence, of F_t) is very close to the optimal. So if we find that δ_2 is large, then this would be due to the truncation of the lag length of $\beta(L)$ in (9). The third part, δ_3, does not need to be orthogonal to δ_2 because the information contained in PC_T is not a part of the information contained in F_T. In fact, the forecast $E^*(y_{T+1}|PC_T)$ may be even better than $E^*(y_{T+1}|F_T)$. This would be the case, for example, if the hypothetical diffusion indexes useful for forecasting y_{T+1} were to partially rely on important sector-specific information, and if PC_T were to absorb such information. Finally, the fourth part of the error, δ_4, is due to replacing the linear forecast $E^*(y_{T+1}|PC_T)$ by a non-linear forecast $\hat{y}_{T+1|T} = \hat{\alpha} + \hat{\beta}'\widehat{PC}_T$. Since $E^*(y_{T+1}|PC_T)$ is not optimal forecast, it is not necessarily true that δ_4 is orthogonal to δ_2 and/or δ_3. In particular, the estimation errors in $\hat{\alpha}, \hat{\beta}$, and $\widehat{PC}_T$ may, in principle, improve the quality of the forecast. Typically, however, we would expect the estimation errors to hurt the forecasting quality, although, according to the asymptotic theory developed by Bai and Ng (2006), such a negative influence would be expected to be negligible for large n and T.

We assess the accuracy of the forecast $\hat{y}_{T+1|T}$ of output growth and of aggregate inflation and the properties of the corresponding error decompositions into $\delta_1, ..., \delta_4$ in the following Monte Carlo experiment. We simulate a large number of datasets of cross-sectional size $n = 156$ and temporal size 120. We use the first 119 values of the simulated series for estimating and the last value for the forecast quality assessment. First, we compute the percentage of the forecastable variance captured by different forecasts. Precisely, we compute

$$\frac{\widehat{Var}(y_{T+1}) - \widehat{Var}(\varepsilon_{T+1})}{\widehat{Var}(y_{T+1}) - \widehat{Var}\left(\varepsilon^o_{T+1}\right)},$$

where ε^o_{T+1} is the error of the optimal forecast $E(y_{T+1}|Y_T, Y_{T-1}, ...)$ and ε_{T+1} is either the error of the (infeasible) forecast $E^*(y_{T+1}|f_T, f_{T-1}, ...)$, or the error of the (infeasible) forecast $E^*(y_{T+1}|F_T)$, or the error of the (infeasible) forecast $E^*(y_{T+1}|PC_T)$, or the error of the (feasible) forecast $\hat{y}_{T+1|T}$. The value of $\widehat{Var}(\varepsilon_{T+1})$ for the latter three forecasts equals the sample Monte Carlo variance (based on 40,000 Monte Carlo replications of the data-generating process) of $\delta_1 + \delta_2$, $\delta_1 + \delta_2 + \delta_3$, and $\delta_1 + \delta_2 + \delta_3 + \delta_4$, respectively. The results of our computations are reported in Table 2.[16]

We see that for output growth, there is little loss from basing the forecast on the current

[16] The figures in the third row of the table, corresponding to the forecast $E^*(y_{T+1}|f_T, f_{T-1}, ...)$, were theoretically obtained and were reported above. We repeat them here for convenience.

Table 2: The percentage of the forecastable variance captured by different forecasts.

Type of forecast	Output growth	Inflation
Optimal	100%	100%
$E^*\left(y_{T+1}\middle\vert f_T, f_{T-1}, ...\right)$	98.8%	97.4%
$E^*\left(y_{T+1}\middle\vert F_T\right)$	90.8%	79.8%
$E^*\left(y_{T+1}\middle\vert PC_T\right)$	92.2%	84.8%
$\hat{y}_{T+1\vert T}$	38.0%	18.3%

and one-period-lagged macroeconomic shocks, which constitute the entries of F_T, relative to the entire history of the macroeconomic shocks $\{f_T, f_{T-1}, f_{T-2}, ...\}$. However, for aggregate inflation, the corresponding loss is larger (see the drop from 97.4% to 79.8% in the third row of Table 2). Next, basing forecast on the first six population principal components of the data rather than on (the six-dimensional vector of the macroeconomic static factors) F_T somewhat improves the forecasting power. Finally, the need for estimating the principal components and the coefficients in the forecast equations lead to very substantial drops in the forecasting power for both output growth and aggregate inflation.

Our Monte Carlo estimates of the covariance matrix of the diffusion index forecast components δ_1, δ_2, δ_3, and δ_4 (based on 40,000 Monte Carlo replications and normalized by the variance of the forecasted series) are as follows:

$$\widehat{Var}\left(\delta\right) = \begin{pmatrix} 0.848 & 0 & 0 & 0 \\ 0 & 0.014 & -0.014 & 0.004 \\ 0 & -0.014 & 0.027 & -0.005 \\ 0 & 0.004 & -0.005 & 0.086 \end{pmatrix}$$

for output growth, and

$$\widehat{Var}\left(\delta\right) = \begin{pmatrix} 0.862 & 0 & 0 & 0 \\ 0 & 0.028 & -0.030 & -0.006 \\ 0 & -0.030 & 0.053 & 0.006 \\ 0 & -0.006 & 0.006 & 0.093 \end{pmatrix}$$

for aggregate inflation. As expected, δ_1, being the error of the optimal forecast, is not correlated to either of δ_2, δ_3, or δ_4. The variance of δ_1 is substantially larger than the variances of δ_2, δ_3, and δ_4, which reflects the fact that the output growth and aggregate inflation series are poorly forecastable in our data. However, the variance of $\delta_2 + \delta_3 + \delta_4$, which can be interpreted as the part of the variance of the diffusion index forecast which is due to its suboptimality is not negligible relative to the variance of δ_1. For output growth, $\widehat{Var}\left(\delta_2 + \delta_3 + \delta_4\right)$ constitutes 11.1% of $\widehat{Var}\left(\delta_1\right)$. For aggregate inflation, $\widehat{Var}\left(\delta_2 + \delta_3 + \delta_4\right)$ constitutes 13.1% of $\widehat{Var}\left(\delta_1\right)$.

8. Structural FAVAR analysis

The availability of large datasets and recent advances in the methodology has shifted the center of empirical macroeconomic analysis from structural Vector Autoregressions (VAR) to structural factor analysis. Factor Augmented Vector Autoregressions (FAVAR) introduced by Bernanke et al. (2005) have been successfully applied to study the effects of monetary policy (Bernanke et al., 2005), the effects of the Euro on the monetary transmission mechanism (Boivin et al., 2008), the extent of price stickiness (Boivin et al., 2009), the sources of global business cycle (Bagliano and Morana, 2009), the economic effects of oil price and credit market shocks (Lescaroux and Mignon, 2009, and Gilchrist et al., 2009), the transmission of international shocks (Mumtaz and Surico, 2009), and the term structure of interest rates (Moench, 2008). An alternative, but closely related, structural factor model approach has been used to analyze the transmission of the common euro-area shocks to new EU member states (Eickmeier and Breitung, 2006), the effects of monetary policy (Forni et al., 2009, and Forni and Gambetti, 2009), and the sectoral-aggregate decomposition of technological shocks (Foester et al, 2008).

The main advantage of the structural factor analysis over the traditional structural VARs is that it uses much richer datasets. Macroeconomic shocks are often non-fundamental relative to a small set of variables utilized by structural VARs. Therefore, the recovered shocks differ from the true ones. The difference results in anomalies such as the price puzzle, the liquidity puzzle, etc. In contrast, macroeconomic shocks are much less likely to be non-fundamental relative to a large set of variables utilized by structural factor analysis. Hence, the true shocks are recovered and the anomalies disappear.[17]

Another advantage of the structural factor analysis is the possibility to analyze the effects of the structural shocks on the disaggregated series. Bernanke et al. (2005) use a FAVAR to estimate the impulse responses of 20 different macroeconomic indicators to a monetary policy shock. Boivin, Giannoni and Mihov (2009) (BGM) use a FAVAR to study the impulse responses of the disaggregated prices to a monetary policy shock. They conclude that the response of the sectoral prices to the monetary policy shock is gradual, so that the macroeconomic component of the sectoral prices is sticky, which is an important finding validating a large body of theoretical research based on the assumption of sticky prices.

In this section, we would like to assess the quality of the FAVAR analysis of the impulse responses of disaggregated macroeconomic series to the monetary policy shock. We will assume that the shock is observable and will include it into the FAVAR as one observable factor, correctly ordered first

[17] See Fernandez-Villaverde et al. (2007) and Forni et al. (2009) for two detailed recent discussions of the non-fundamentalness issue.

in the causal chain relating it to the other (unobserved) factors. This way, we abstract from the problems that may be caused by the incorrect identification of the shock, and focus only on the problems specifically related to the fact that the factor model is an imperfect representation of the underlying DSGE model. We start by estimating five latent factors in our simulated data, assuming that the sixth factor is the monetary policy shock, which we assume to be observed. We choose to consider five latent factors to make our FAVAR analysis similar to that in BGM. We will also analyze the changes which would result from different assumptions about the number of factors. To extract the latent factors from our data we use the same methodology as BGM. Precisely, first, we obtain the initial estimates of the five factors using the principal components method. Then, we regress our data on the so extracted factors and on the monetary policy shock. We compute an auxiliary dataset, which is our initial data net of the estimated monetary policy shock component, and extract five factors from such an auxiliary dataset. Then, we again regress our initial data on the newly extracted factors and on the monetary policy shock. After that, we compute an auxiliary data set, which is our initial data net of the estimated monetary policy shock component and so on. We make 20 iterations (more than enough for the convergence) and take the factors and loadings estimated in the last iteration as our final estimates.

8.1 Effects of monetary policy shocks

We proceed with a description of the response of our data series to a shock to money growth. As explained above, we treat the monetary policy shock as an observable correctly ordered in the causal chain, so there is no issue of inaccurate recovery of the monetary policy shock by a VAR. In particular, our analysis below does not attempt to explain the workings of the FAVAR solving puzzles such as the price puzzle. We are only concerned with the accuracy of the FAVAR estimates of the impulse responses of disaggregated series to the correctly identified monetary policy shock.

Figure 16 shows the impulse responses of aggregate consumption, aggregate inflation, aggregate output and aggregate hours to a one standard deviation increase in the rate of money growth. We identify the shock by ranking money growth first in the causation chain. That is, in contrast to BGM, the money growth rate is assumed not to be contemporaneously affected by any other variables because, in the model, money growth is an exogenous process.

The dashed lines in Figure 16 correspond to the medians, 5% and 95% percentiles of the Monte Carlo distribution (we make 1000 Monte Carlo replications) of the FAVAR estimates of the impulse responses. The solid lines corresponds to the theoretical impulse responses. We see that the estimated impulse responses are similar to the theoretical ones. However, on average, they decay somewhat faster than the theoretical impulse responses.

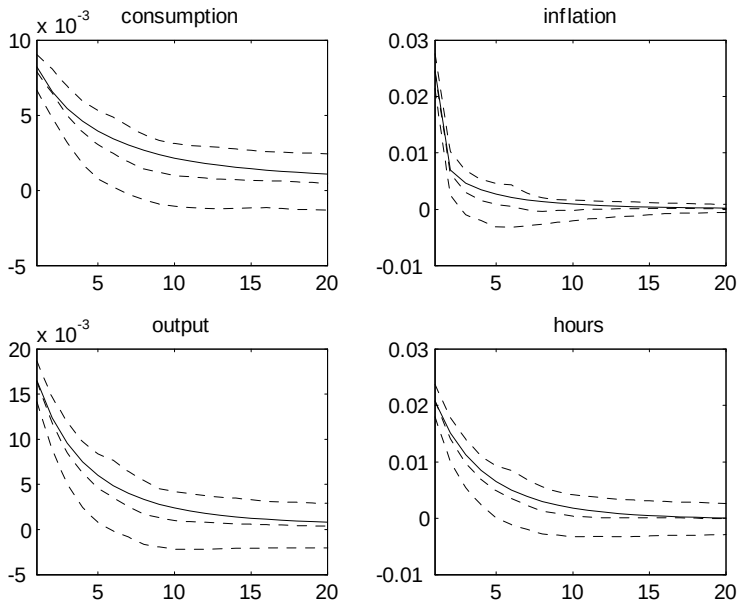

Figure 16: Impulse responses of different aggregate variables to a one standard deviation shock to money growth innovation. Dashed lines: 5, 50 and 90 percentiles of the monte carlo distribution of estimated impulse responses. Solid lines: theoretical impulse responses.

Figures 17, 18, 19, 20 and 21 report the impulse responses of sector-specific consumptions, inflations, outputs, hours, and wages, respectively, to a shock of one standard deviation in the rate of money growth. The estimated impulse responses remain similar to the theoretical ones. However, the uncertainty caused by the estimation becomes large for some variables and sectors.

8.2 Different number of factors

Here we repeat the above analysis using a different number of latent factors in the FAVAR. Figure 22 shows the impulse responses of the aggregate variables to the monetary policy shock when the number of latent factors is only 1. Somewhat surprisingly, the figure looks very similar to Figure 16, which is based on five latent factors. To facilitate the comparison, we show the impulse responses percentiles from Figure 16 as dotted lines on Figure 22.

The entire Monte Carlo distribution of the impulse responses estimated using a FAVAR with one latent factor is somewhat shifted upward relative to the distribution of the impulse responses estimated using a FAVAR with five latent factors. This shift is much more noticeable at longer horizons. On average, the Monte Carlo distribution of the impulse responses based on the 5-factor FAVAR has somewhat lower variance than that of the responses based on a 1-factor FAVAR. However, overall, the differences are small.

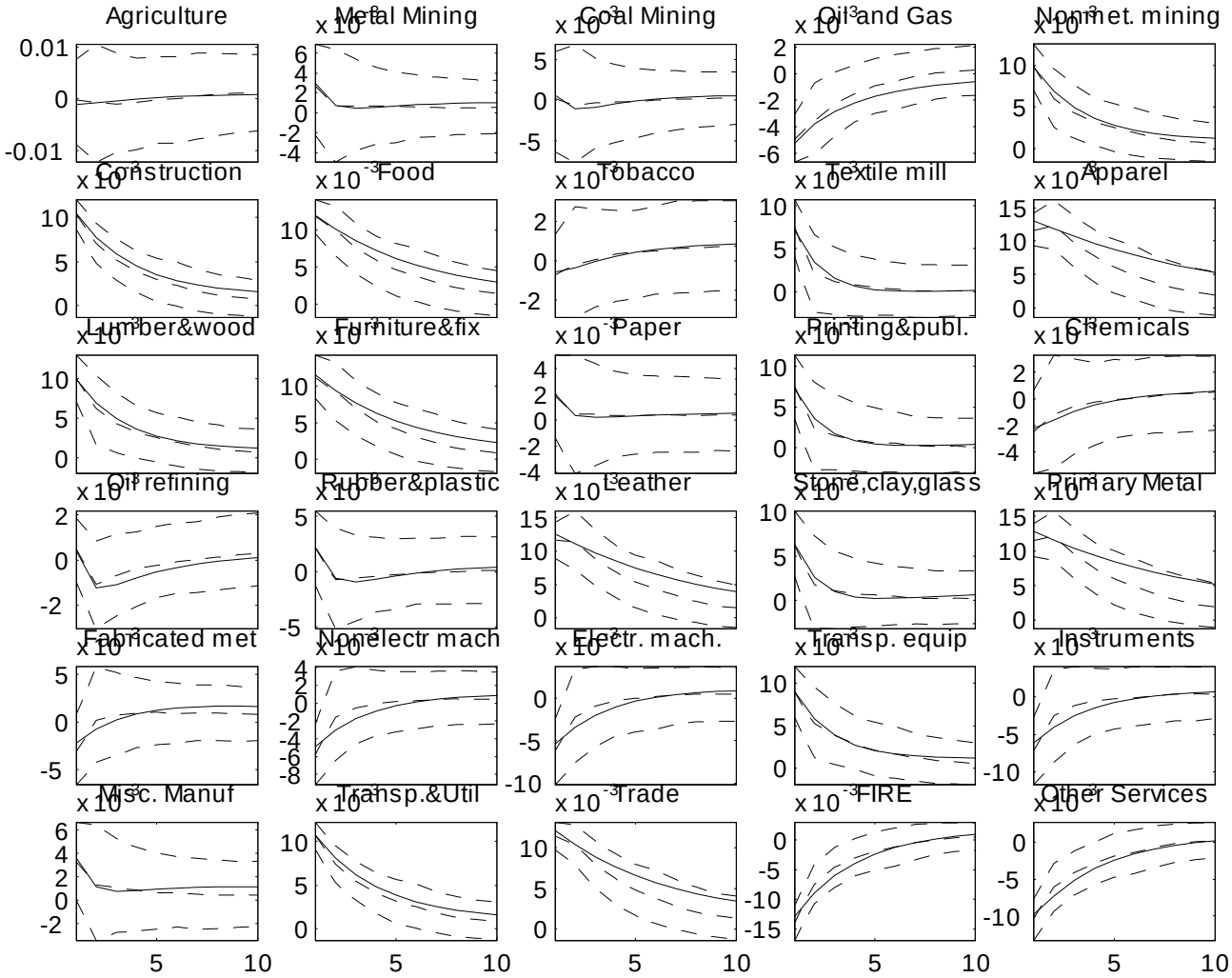

Figure 17: Estimated and theoretical impulse responses of sectoral consumptions to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the estimated impulse responses.

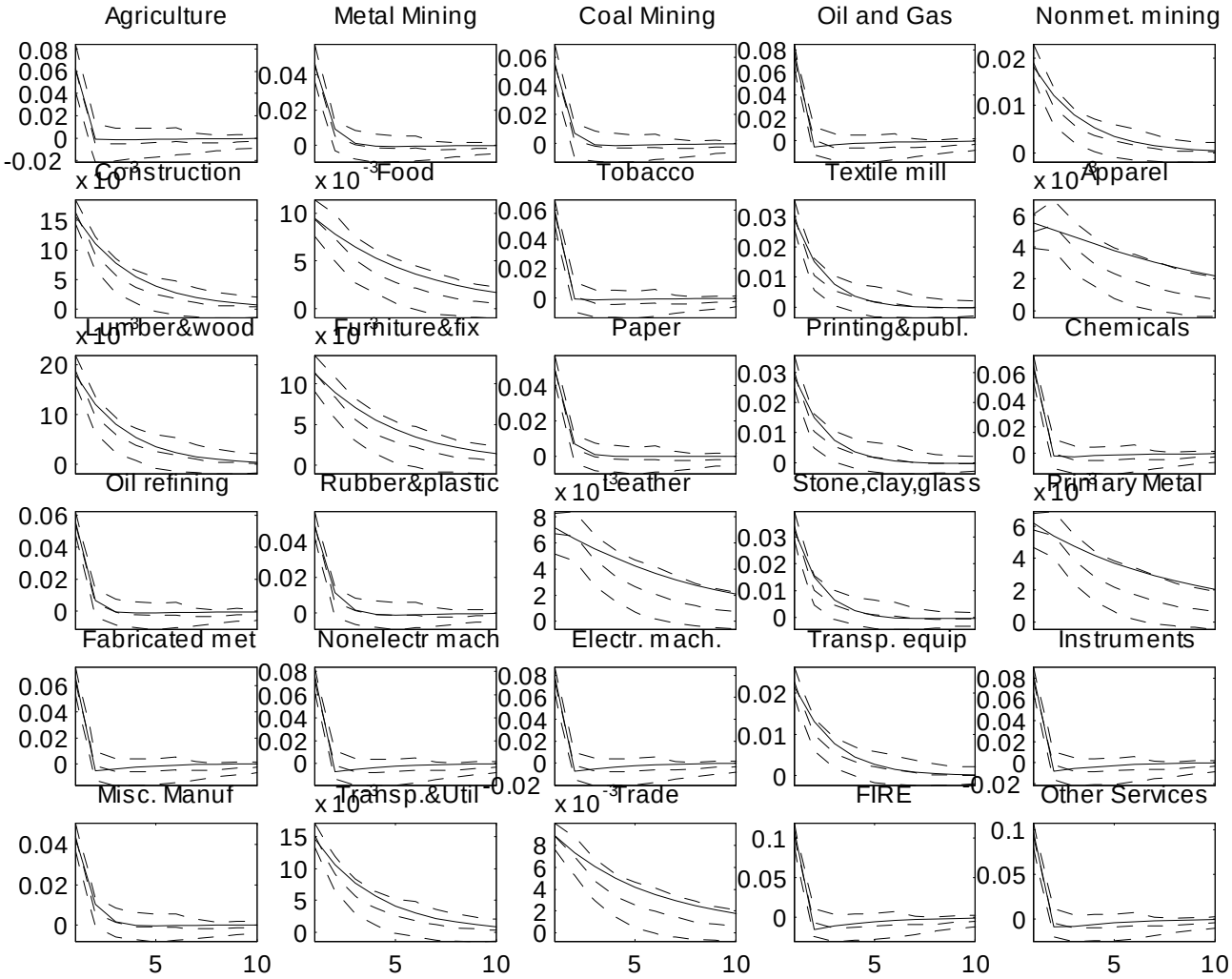

Figure 18: Estimated and theoretical impulse responses of sectoral inflations to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the estimated impulse responses.

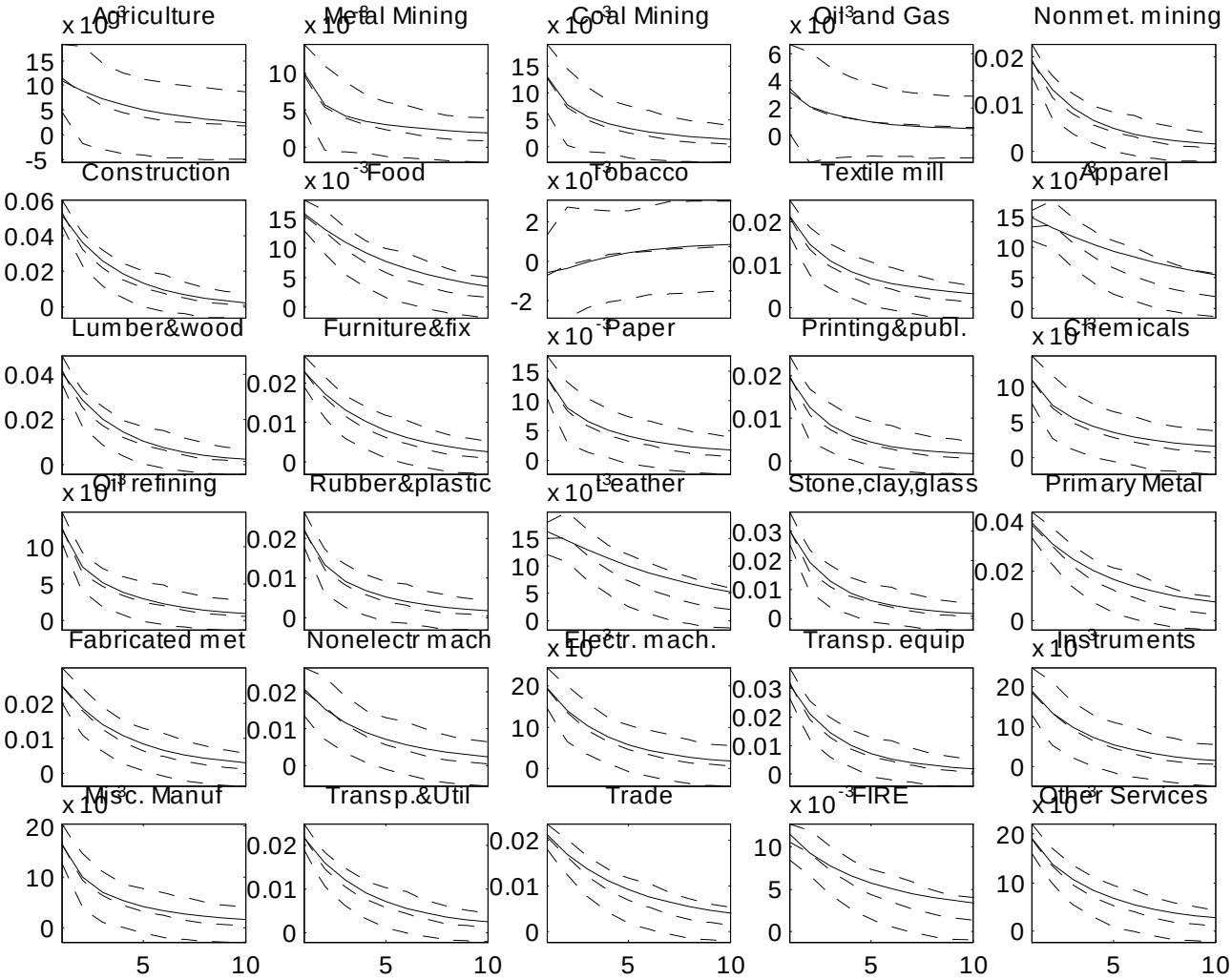

Figure 19: Estimated and theoretical impulse responses of sectoral outputs to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the estimated impulse responses.

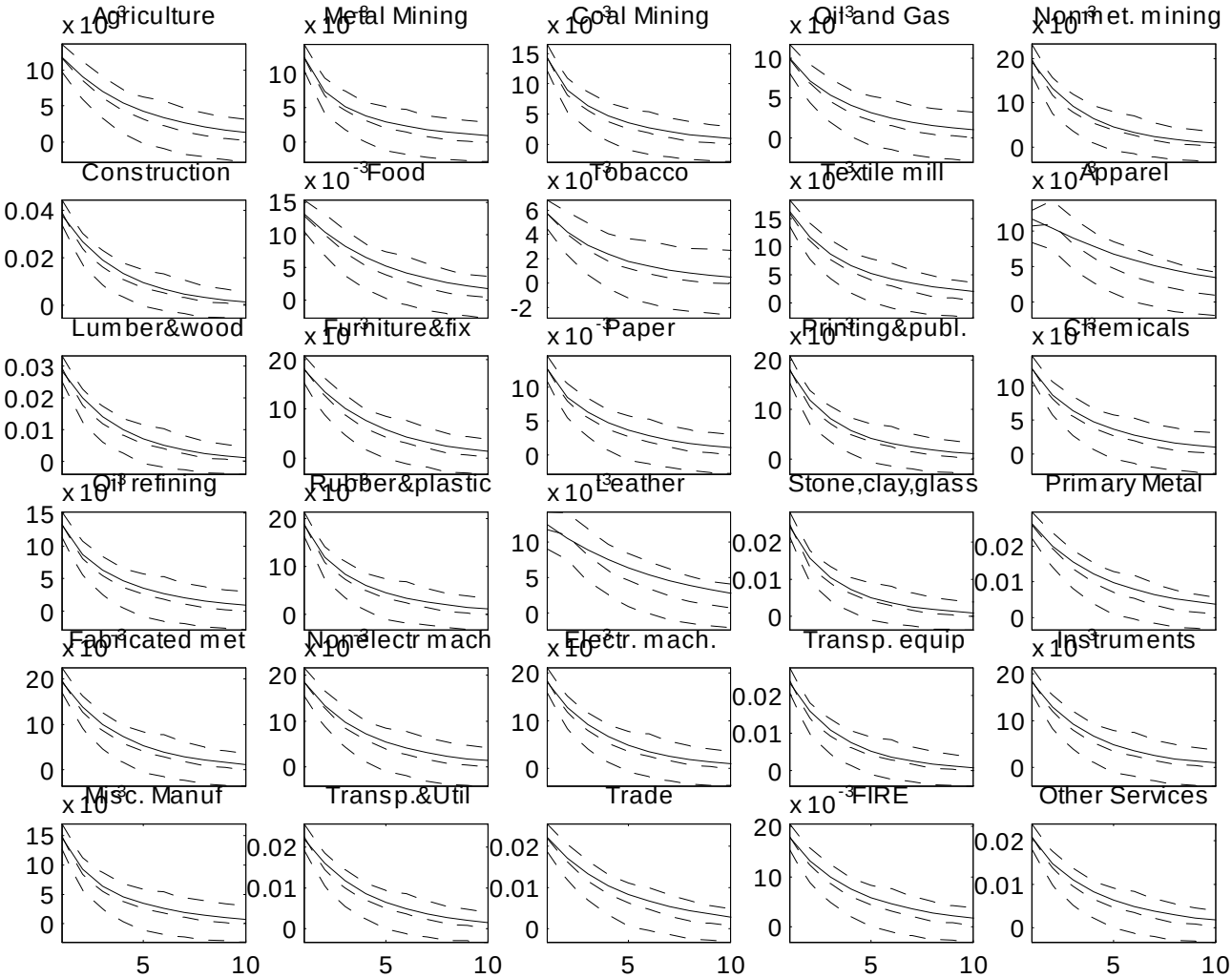

Figure 20: Estimated and theoretical impulse responses of sectoral hours to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the estimated impulse responses.

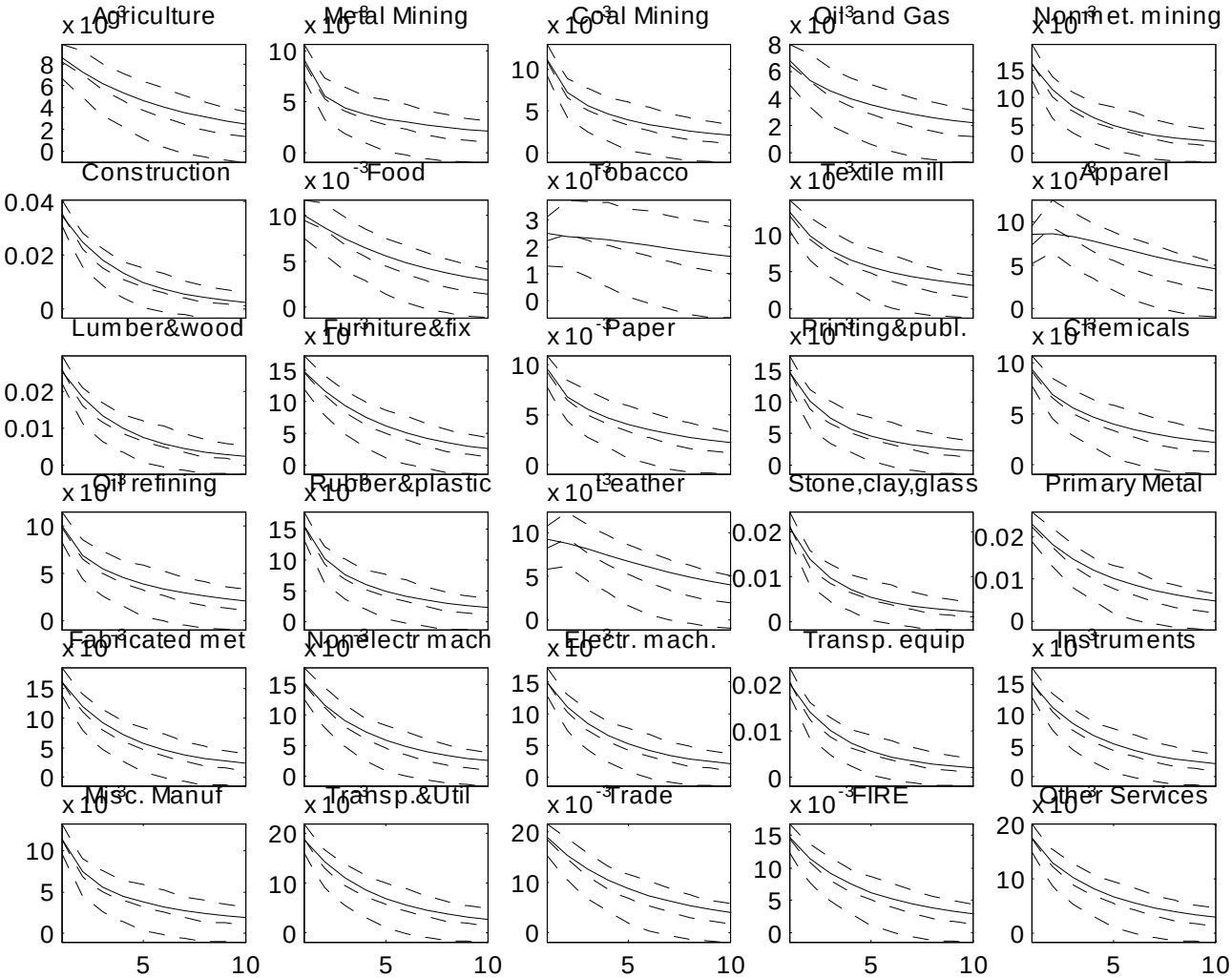

Figure 21: Estimated and theoretical impulse responses of sectoral wages to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the estimated impulse responses.

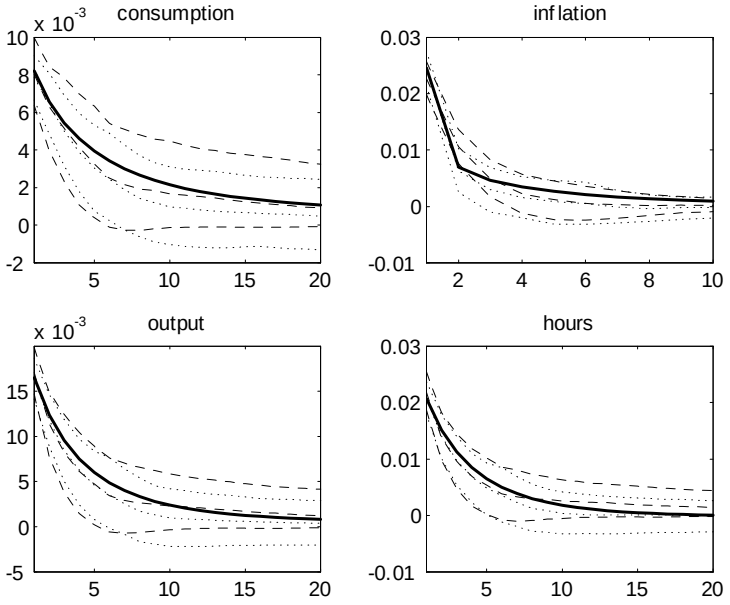

Figure 22: Impulse responses of different aggregate variables to a one standard deviation shock to money growth innovation. Dashed lines: 5, 50 and 90 percentiles of the monte carlo distribution of impulse responses estimated using FAVAR with 1 latent factor. Solid lines: theoretical impulse responses. Dotted lines: 5, 50 and 90 percentiles of the monte carlo distribution of impulse responses estimated using FAVAR with 5 latent factors.

Figures 23, 24, 25, 26 and 27 superimpose the impulse responses of sectoral consumptions, inflations, outputs, hours and wages estimated by a FAVAR with 1 factor (dashed lines) and by a FAVAR with five factors (dotted lines). We see that many sectoral impulse responses are estimated much better by a FAVAR with 5 factors than by a FAVAR with 1 factor.

Table 3 shows how the accuracy of the impulse response estimation depends on the number of factors in more detail. For each sector and each particular category of variables, we compute the mean squared error of the impulse response estimate based on FAVARs with 1 through 6 latent factors. The mean squared error is cumulative over the first 8 periods of the response. Its computation is based on 1,000 Monte Carlo replications. We divide all the obtained mean square errors by the corresponding mean squared error of the estimate from a FAVAR with 1 factor. Then, for each category of variables, we report the average and standard deviations (in parentheses) of these ratios over 30 different sectors.

Table 3: Means and standard deviations (over 30 sectors) of the Monte Carlo mean squared errors of the impulse responses estimated by FAVARs using different number of latent factors. The mean squared errors are cumulative over the first 8 periods. They are normalized so that the mean squared error of the estimates obtained from a FAVAR based on a single factor equals one.

Number of factors	Consumption	Inflation	Output	Hours	Wages
1	1	1	1	1	1
2	0.73 (0.16)	0.89 (0.33)	0.75 (0.19)	0.80 (0.13)	0.24 (0.04)
3	0.80 (0.23)	0.70 (0.28)	0.85 (0.26)	0.88 (0.14)	0.27 (0.05)
4	0.69 (0.23)	0.54 (0.41)	0.88 (0.27)	0.90 (0.13)	0.28 (0.06)
5	0.61 (0.20)	0.55 (0.44)	0.84 (0.24)	0.91 (0.12)	0.29 (0.06)
6	0.58 (0.20)	0.55 (0.44)	0.83 (0.24)	0.93 (0.13)	0.29 (0.07)

We see that using two latent factors instead of one substantially improves the quality of the FAVAR estimates of the impulse responses of variables in all the categories. The improvement is especially dramatic for sectoral wages. Further increasing the number of factors gives mixed results. For output, hours and wages, the quality of the estimated impulse responses deteriorates, while for consumption and inflation, it further improves. This finding is consistent with the finding above that the information about the true macroeconomic factors is spread through a relatively large number of principal components even though the main portion of the macroeconomic factors' variation is captured by the first two or three principal components.

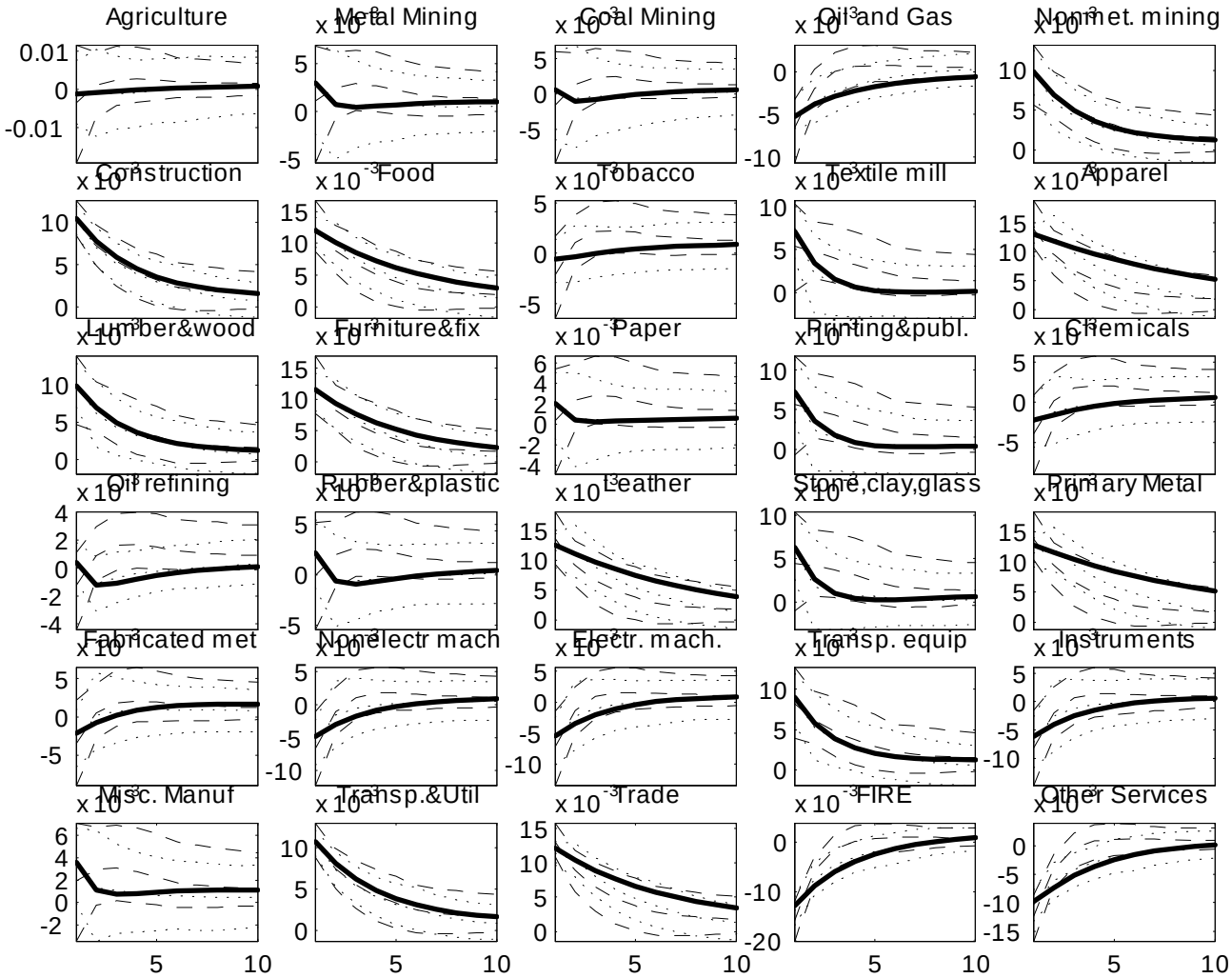

Figure 23: Estimated and theoretical impulse responses of sectoral consumptions to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the impulse responses estimated by FAVAR with 1 factor. The dotted line correspond to the estimation by FAVAR with 5 factors.

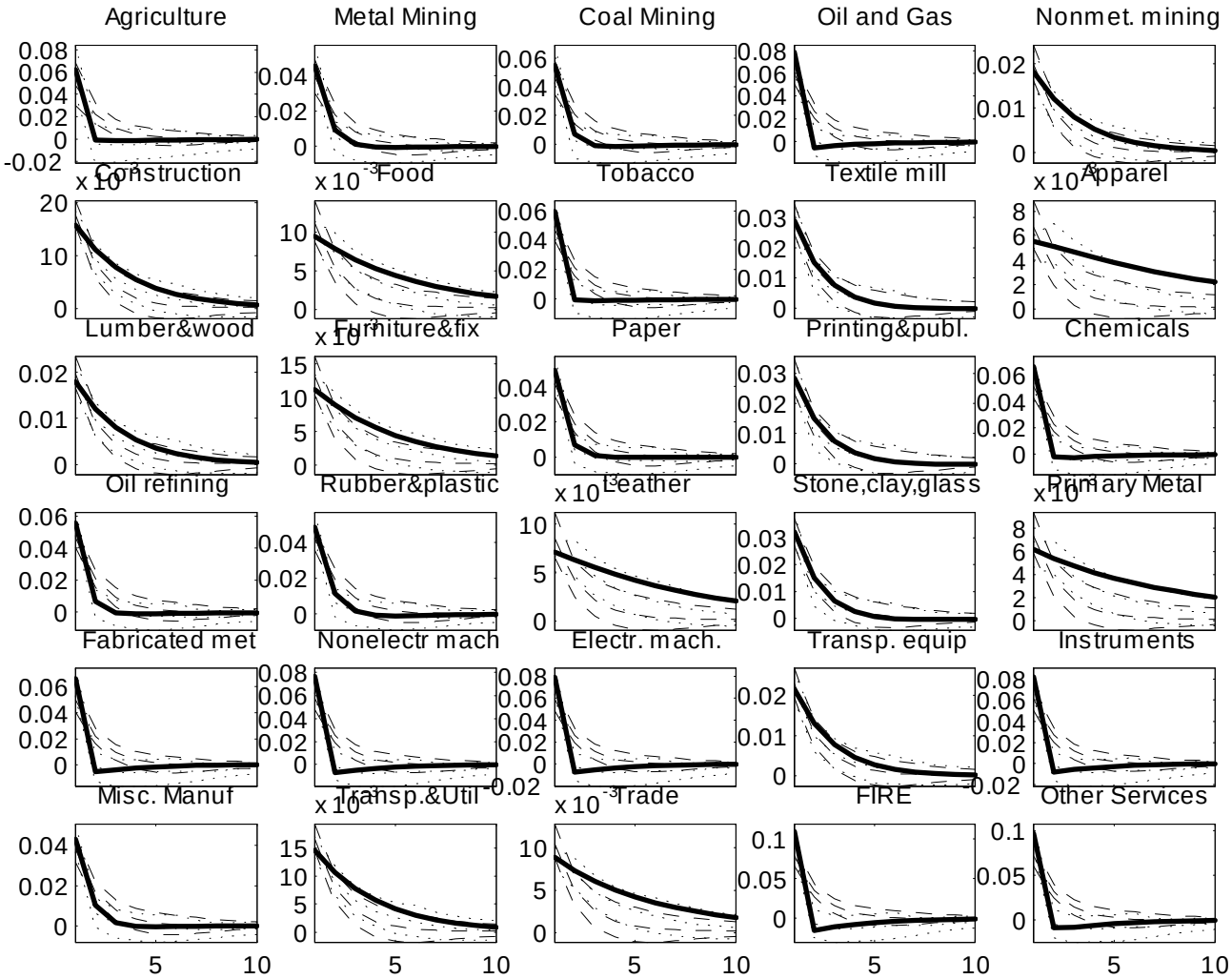

Figure 24: Estimated and theoretical impulse responses of sectoral inflations to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the impulse responses estimated by FAVAR with 1 factor. The dotted line correspond to the estimation by FAVAR with 5 factors.

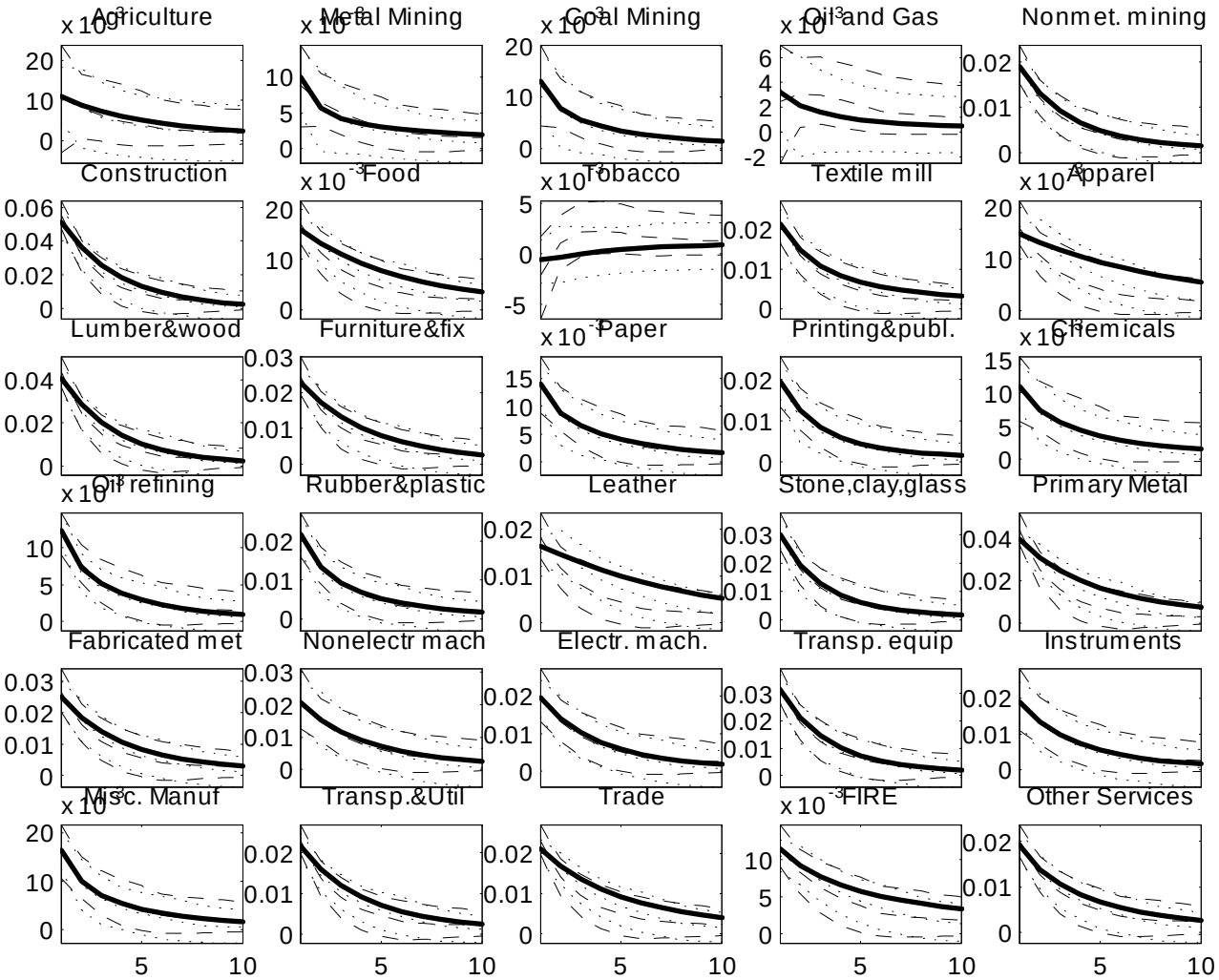

Figure 25: Estimated and theoretical impulse responses of sectoral outputs to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the impulse responses estimated by FAVAR with 1 factor. The dotted line correspond to the estimation by FAVAR with 5 factors.

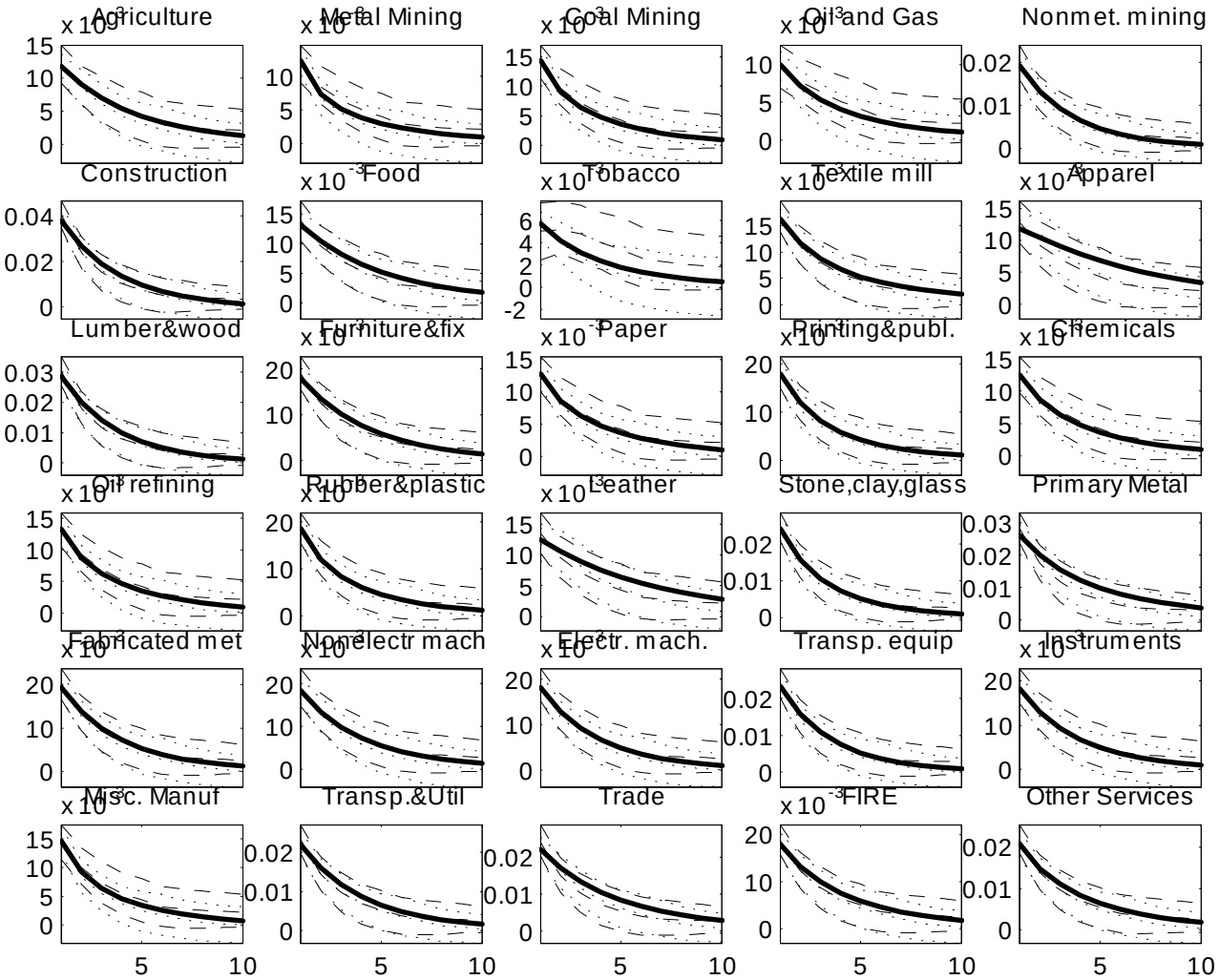

Figure 26: Estimated and theoretical impulse responses of sectoral hours to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the impulse responses estimated by FAVAR with 1 factor. The dotted line correspond to the estimation by FAVAR with 5 factors.

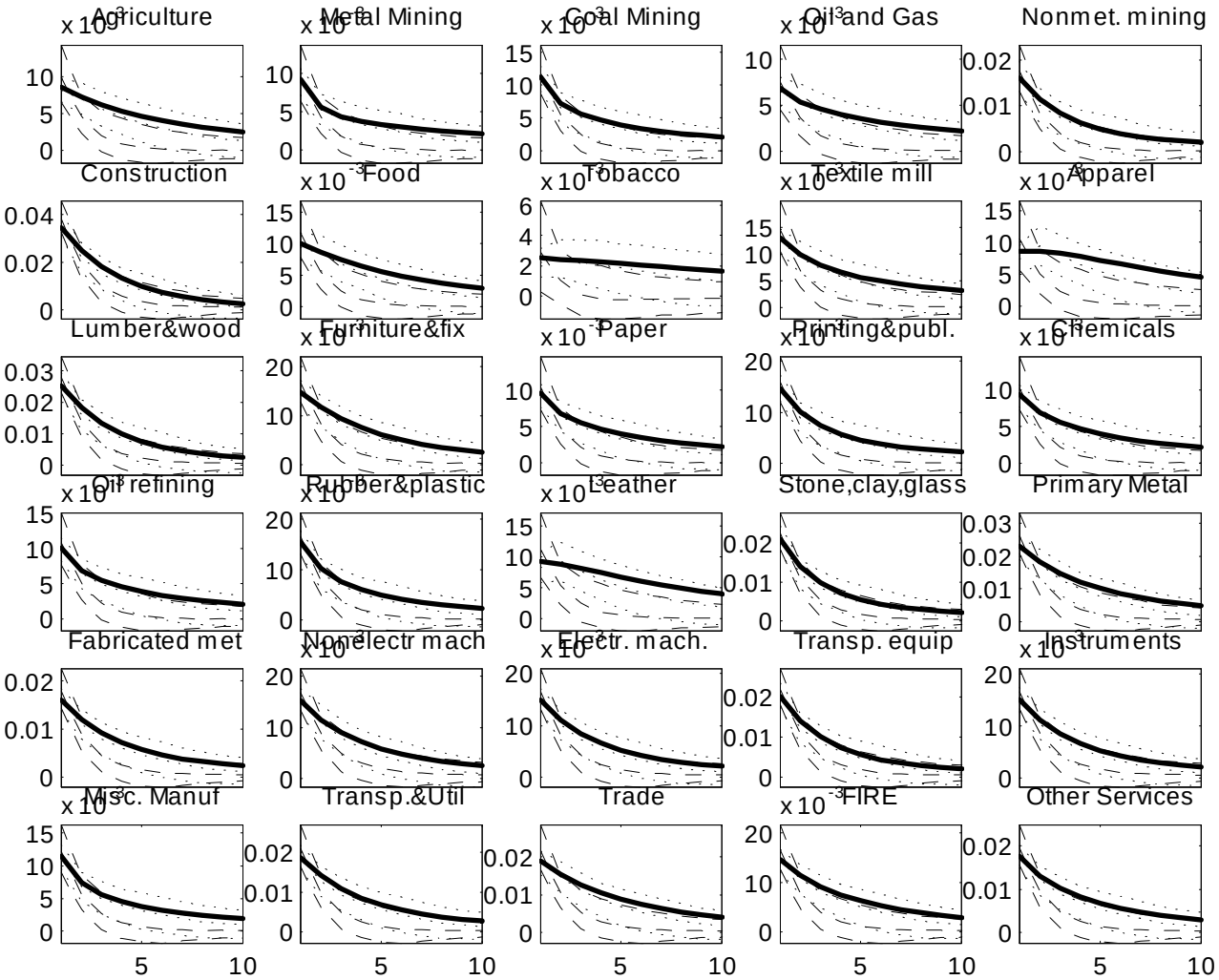

Figure 27: Estimated and theoretical impulse responses of sectoral wages to monetary policy shock. The dashed lines correspond to 5, 50 and 95 percentiles of the Monte Carlo distribution of the impulse responses estimated by FAVAR with 1 factor. The dotted line correspond to the estimation by FAVAR with 5 factors.

9. Conclusions

In this paper, we have used a highly-dissagregated, multi-sector DSGE model as a laboratory to shed some light on the application of factor analysis to economic data. As in actual applications, we have access to a large number of disaggregated series with rich dynamics and face a certain ambiguity as to what "factors" are. However, in contrast to actual applications, we know the true data generating process and this knowledge allow us to explore the macroeconomic content of factor analysis and to assess the practical benefits and limitations of applying these statistical tools to real-world data.

We find that among the three aggregate and thirty sectoral shocks in the model, only the former may be thought of as factors in the sense that they non-trivially affect most of the 156 variables in the dataset. This result supports the view, implicit in most of the applied research, that the factor space may be associated with the space of basic macroeconomic shocks. However, despite the pervasiveness of the aggregate shocks, the principal components analysis has a difficult time replicating the macroeconomic factor space. Also, the asymptotic approximation to the distribution of the principal components is relatively poor when applied to our sample size, which, as it happens, is of the length typically used in most applied studies. The application of standard procedures to determine the number of factors deliver different results depending on the criteria used and auxiliary parameters. Development of sensible procedures for assessing the relative quality of the critera in the situation when the concept of the true factor is not well-defined, as in our application, is left for future research.

On the other hand, we find that diffusion index forecasting performs reasonably well on our simulated data and that factor augmented vector autoregressions (FAVAR) accurately recover the true impulse responses to a monetary policy shock at least when it is treated as an observable correctly ordered in the causal chain. These two results provide support the use of factor analysis techniques for forecasting, nowcasting, and policy analysis.

A The Multi-Sector Model

This Appendix describes the multi-sector model and parameter values used to generate the artificial data used in our analysis. The model is that developed by Bouakez, Cardia and Ruge-Murcia (2009) and we follow closely their presentation. For additional details on the model and its econometric estimation, we refer the reader to their article.

1.1 Production and Intermediate Consumption

Output is produced in J heterogenous sectors. In each sector there is a continuum of monopolistically-competitive firms that produce a differentiated good but are identical otherwise. Firms in different sectors face different nominal frictions, and use different production functions, and combinations of material and investment inputs. These assumptions imply that the equilibrium will be symmetric within sectors but asymmetric between sectors.

Firm l in sector j produces output y_t^{lj} using the technology

$$y_t^{lj} = (z_t^j n_t^{lj})^{\nu^j} (k_t^{lj})^{\alpha^j} (H_t^{lj})^{\gamma^j}, \tag{15}$$

where z_t^j is a sector-specific productivity shock, n_t^{lj} is labor, k_t^{lj} is capital, H_t^{lj} is materials inputs, and $\nu^j, \alpha^j, \gamma^j \in (0,1)$ and satisfy $\nu^j + \alpha^j + \gamma^j = 1$. The sectoral productivity shock follows the process

$$\ln(z_t^j) = (1 - \rho_{z^j}) \ln(z_{ss}^j) + \rho_{z^j} \ln(z_{t-1}^j) + \epsilon_{z^j,t},$$

where $\rho_{z^j} \in (-1,1)$, $\ln(z_{ss}^j)$ is the unconditional mean, and the innovation $\epsilon_{z^j,t}$ is identically and independently distributed (*i.i.d.*) with zero mean and variance $\sigma_{z^j}^2$.

Materials inputs are an aggregate of goods produced by all firms in all sectors. In particular,

$$H_t^{lj} = \prod_{i=1}^{J} \zeta_{ij}^{-\zeta_{ij}} (h_{i,t}^{lj})^{\zeta_{ij}}, \tag{16}$$

where

$$h_{i,t}^{lj} = \left(\int_0^1 \left(h_{mi,t}^{lj} \right)^{(\theta-1)/\theta} dm \right)^{\theta/(\theta-1)}, \tag{17}$$

$h_{mi,t}^{lj}$ is the quantity of materials purchased from firm m in sector i, $\zeta_{ij} \in (0,1)$ is a weight that satisfies $\sum_{i=1}^{J} \zeta_{ij} = 1$, and $\theta > 1$ is the elasticity of substitution between goods produced in the same sector.

The capital stock is directly owned by the firm and evolves according to

$$k_{t+1}^{lj} = (1 - \delta) k_t^{lj} + X_t^{lj}, \tag{18}$$

where $\delta \in (0,1)$ is the rate of depreciation and X_t^{lj} denotes an investment technology that aggregates different investment goods into units of capital. In particular,

$$X_t^{lj} = \prod_{i=1}^{J} \kappa_{ij}^{-\kappa_{ij}} (x_{i,t}^{lj})^{\kappa_{ij}}, \tag{19}$$

where

$$x_{i,t}^{lj} = \left(\int_0^1 \left(x_{mi,t}^{lj} \right)^{(\theta-1)/\theta} dm \right)^{\theta/(\theta-1)}, \tag{20}$$

$x_{mi,t}^{lj}$ is the quantity of investment goods purchased from firm m in sector i, and $\kappa_{ij} \in (0,1)$ is a weight that satisfies $\sum_{i=1}^{J} \kappa_{ij} = 1$. The prices of the composites H_t^j and X_t^j are, respectively, $Q_t^{H^j} = \prod_{i=1}^{J} (p_t^i)^{\zeta_{ij}}$ and $Q_t^{X^j} = \prod_{i=1}^{J} (p_t^i)^{\kappa_{ij}}$, where

$$p_t^i = \left(\int_0^1 (p_t^{mi})^{1-\theta} dm \right)^{1/(1-\theta)} \tag{21}$$

and p_t^{mi} is the price of the good produced by firm m in sector i.

Firms face convex costs when adjusting their capital stock and nominal prices. The capital-adjustment cost takes the form

$$\Gamma_t^{lj} = \Gamma(X_t^{lj}, k_t^{lj}) = \frac{\chi}{2} \left(\frac{X_t^{lj}}{k_t^{lj}} - \delta \right)^2 k_t^{lj}, \tag{22}$$

where $\chi > 0$. The real per-unit cost of changing the nominal price is

$$\Phi_t^{lj} = \Phi(p_t^{lj}, p_{t-1}^{lj}) = \frac{\phi^j}{2} \left(\frac{p_t^{lj}}{\pi_{ss} p_{t-1}^{lj}} - 1 \right)^2, \tag{23}$$

where p_t^{lj} is the price of the good produced by firm l in sector j, π_{ss} is the steady-state aggregate inflation rate and $\phi^j \geqslant 0$ is a sector-specific parameter.

The firm's problem is to maximize

$$E_\tau \sum_{t=\tau}^{\infty} \beta^{t-\tau} \left(\frac{\Lambda_\tau}{\Lambda_t} \right) \left(\frac{d_t^{lj}}{P_t} \right), \tag{24}$$

where d_t^{lj} are nominal profits, P_t is the aggregate price index (see below), $\beta \in (0,1)$ is a discount

factor and Λ_t is the consumers' marginal utility of wealth. Nominal profits are

$$\begin{aligned} d_t^{lj} = \; & p_t^{lj}\left(c_t^{lj} + \sum_{i=1}^{J}\int_0^1 x_{lj,t}^{mi}dm + \sum_{i=1}^{J}\int_0^1 h_{lj,t}^{mi}dm\right) - w_t^{lj}n_t^{lj} - \sum_{i=1}^{J}\int_0^1 p_t^{mi}x_{mi,t}^{lj}dm - \sum_{i=1}^{J}\int_0^1 p_t^{mi}h_{mi,t}^{lj}dm \\ & -\Gamma_t^{lj}Q_t^{X^j} - \Phi_t^{lj}p_t^{lj}\left(c_t^{lj} + \sum_{i=1}^{J}\int_0^1 x_{lj,t}^{mi}dm + \sum_{i=1}^{J}\int_0^1 h_{lj,t}^{mi}dm\right), \end{aligned} \tag{25}$$

where c_t^{lj} is final consumption, w_t^{lj} is the nominal wage, and $x_{lj,t}^{mi}$ and $h_{lj,t}^{mi}$ are respectively the quantities sold to firm m in sector i as materials input and investment good. The solution of the firm's problem delivers optimal demand functions for materials and investment inputs:

$$\begin{aligned} x_{mi,t}^{lj} &= \kappa_{ij}\left(p_t^{mi}/p_t^i\right)^{-\theta}\left(p_t^i/Q_t^{X^j}\right)^{-1}X_t^{lj}, \\ h_{mi,t}^{lj} &= \zeta_{ij}\left(p_t^{mi}/p_t^i\right)^{-\theta}\left(p_t^i/Q_t^{H^j}\right)^{-1}H_t^{lj}. \end{aligned}$$

For these demand functions, $\sum_{i=1}^{J}\int_0^1 p_t^{mi}x_{mi,t}^{lj}dm = \sum_{i=1}^{J} p_t^i x_{i,t}^{lj} = Q_t^{X^j}X_t^{lj}$ and $\sum_{i=1}^{J}\int_0^1 p_t^{mi}h_{mi,t}^{lj}dm = \sum_{i=1}^{J} p_t^i h_{i,t}^{lj} = Q_t^{H^j}H_t^{lj}$.

1.2 Final Consumption

Consumers are identical, infinitely lived, and their number is constant and normalized to one. The representative consumer maximizes

$$E_\tau \sum_{t=\tau}^{\infty}\beta^{t-\tau}\left(\log(C_t) + \upsilon_t\log(M_t/P_t) + \eta_t\log(1-N_t)\right), \tag{26}$$

where C_t is consumption, M_t is the nominal money stock, N_t is hours worked, and υ_t and η_t are preference shocks. These shocks follow the processes

$$\begin{aligned} \ln(\upsilon_t) &= (1-\rho_\upsilon)\ln(\upsilon_{ss}) + \rho_\eta\ln(\upsilon_{t-1}) + \epsilon_{\upsilon,t}, \\ \ln(\eta_t) &= (1-\rho_\eta)\ln(\eta_{ss}) + \rho_\eta\ln(\eta_{t-1}) + \epsilon_{\eta,t}, \end{aligned}$$

where $\rho_\upsilon, \rho_\eta \in (-1,1)$, $\ln(\upsilon_{ss})$ and $\ln(\eta_{ss})$ are unconditional means, and the innovations $\epsilon_{\upsilon,t}$ and $\epsilon_{\eta,t}$ are *i.i.d.* with zero mean and variances σ_υ^2 and σ_η^2, respectively.

Consumption is an aggregate of all available goods:

$$C_t = \prod_{j=1}^{J}(\xi^j)^{-\xi^j}(c_t^j)^{\xi^j}, \tag{27}$$

where $\xi^j \in (0,1)$ and satisfies $\sum_{j=1}^{J} \xi^j = 1$, and

$$c_t^j = \left(\int_0^1 \left(c_t^{lj} \right)^{(\theta-1)/\theta} dl \right)^{\theta/(\theta-1)}, \tag{28}$$

with c_t^{lj} the final consumption of the good produced by firm l in sector j. Hours worked are an aggregate of the hours supplied to each firm in each sector:

$$N_t = \left(\sum_{j=1}^{J} (n_t^j)^{(\varsigma+1)/\varsigma} \right)^{\varsigma/(\varsigma+1)}, \tag{29}$$

where $\varsigma > 0$ and $n_t^j = \int_0^1 n_t^{lj} dl$ is the number of hours worked in sector j, with n_t^{lj} being the number of hours worked in firm l in sector j. The aggregate price index is defined as

$$P_t = \prod_{j=1}^{J} (p_t^j)^{\xi^j}, \tag{30}$$

where

$$p_t^j = \left(\int_0^1 (p_t^{lj})^{1-\theta} dl \right)^{1/(1-\theta)}. \tag{31}$$

The consumer's dynamic budget constraint (in real terms) is

$$\begin{aligned} \sum_{j=1}^{J} \int_0^1 \left(\frac{p_t^{lj} c_t^{lj}}{P_t} \right) dl + b_t + m_t + \sum_{j=1}^{J} \int_0^1 \left(\frac{a_t^{lj} s_t^{lj}}{P_t} \right) dl &= \sum_{j=1}^{J} \int_0^1 \left(\frac{w_t^{lj} n_t^{lj}}{P_t} \right) dl + \frac{R_{t-1} b_{t-1}}{\pi_t} + \frac{m_{t-1}}{\pi_t} \\ &\quad + \sum_{j=1}^{J} \int_0^1 \left(\frac{(d_t^{lj} + a_t^{lj}) s_{t-1}^{lj}}{P_t} \right) dl + \frac{\Upsilon_t}{P_t}, \end{aligned}$$

where $b_t = B_t/P_t$ is the real value of nominal bond holdings, $m_t = M_t/P_t$ is real money balances, s_{t-1}^j are shares in a mutual fund $j = 1, \ldots, J$, R_t is the gross nominal interest rate on bonds that mature at time $t+1$, π_t is the gross inflation rate between periods $t-1$ and t, Υ_t is a government lump-sum transfer, and a_t^j and d_t^j are, respectively, the price of a share in, and the dividend paid by, the mutual fund j.

Utility maximization delivers the optimal demand function

$$c_t^{lj} = \xi^j \left(\frac{p_t^{lj}}{p_t^j} \right)^{-\theta} \left(\frac{p_t^j}{P_t} \right)^{-1} C_t. \tag{32}$$

Using this demand function and the definition of the price indices, it is easy to show that $\sum_{j=1}^{J}\int_0^1 p_t^{lj}c_t^{lj}dl = \sum_{j=1}^{J} p_t^j c_t^j = P_t C_t$.

1.3 Monetary Policy

Money is supplied by the government according to $M_t = \mu_t M_{t-1}$, where μ_t is the stochastic gross rate of money growth. The rate of money growth follows the process

$$\ln(\mu_t) = (1-\rho_\mu)\ln(\mu_{ss}) + \rho_\mu \ln(\mu_{t-1}) + \epsilon_{\mu,t},$$

where $\rho_\mu \in (-1,1)$, $\ln(\mu_{ss})$ is the unconditional mean, and the innovation $\epsilon_{\mu,t}$ is *i.i.d.* with zero mean and variance σ_μ^2. Monetary injections are transferred to consumers lump-sum so that the budget constraint

$$\Upsilon_t/P_t = m_t - m_{t-1}/\pi_t \tag{33}$$

is always satisfied.

1.4 Aggregation

In equilibrium, net private bond holdings equal zero because consumers are identical, the total share holdings in sector j add up to one, and firms in the same sector are identical. Hence, the private sector's budget constraint is

$$\sum_{j=1}^{J}\frac{p_t^j c_t^j}{P_t} + m_t = \sum_{j=1}^{J}\frac{w_t^j n_t^j}{P_t} + \sum_{j=1}^{J}\frac{d_t^j}{P_t} + \frac{m_{t-1}}{\pi_t} + \frac{\Upsilon_t}{P_t}. \tag{34}$$

Substituting in the government budget constraint (33) and multiplying through by the price level yield

$$\sum_{j=1}^{J} p_t^j c_t^j = \sum_{j=1}^{J} w_t^j n_t^j + \sum_{j=1}^{J} d_t^j. \tag{35}$$

Define the value of gross output produced by sector j

$$V_t^j \equiv p_t^j\left(c_t^j + \sum_{i=1}^{J} x_{j,t}^i + \sum_{i=1}^{J} h_{j,t}^i\right), \tag{36}$$

and the sum of all adjustment costs in sector j

$$A_t^j = \Gamma_t^j Q_t^{X^j} + \Phi_t^j p_t^j\left(c_t^j + \sum_{i=1}^{J} x_{j,t}^i + \sum_{i=1}^{J} h_{j,t}^i\right). \tag{37}$$

Then, aggregate nominal dividends are

$$\sum_{j=1}^{J} d_t^j = \sum_{j=1}^{J} V_t^j - \sum_{j=1}^{J} w_t^j n_t^j - \sum_{j=1}^{J} Q_t^{X^j} X_t^j - \sum_{j=1}^{J} Q_t^{H^j} H_t^j - \sum_{j=1}^{J} A_t^j. \tag{38}$$

The nominal value added in sector j is denoted by Y_t^j and is defined as the value of gross output produced by that sector minus the cost of materials inputs

$$Y_t^j = V_t^j - Q_t^{H^j} H_t^j. \tag{39}$$

Substituting (38) and (39) into (35), using $\sum_{j=1}^{J} p_t^j c_t^j = P_t C_t$, and rearranging yield

$$\sum_{j=1}^{J} Y_t^j = P_t C_t + \sum_{j=1}^{J} Q_t^{X^j} X_t^j + \sum_{j=1}^{J} A_t^j. \tag{40}$$

That is, aggregate output equals private consumption plus investment and the sum of all adjustment costs in all sectors.

1.5 Parameter Values

Bouakez, Cardia and Ruge-Murcia (2009) consider thirty sectors that roughly correspond to the two-digit Standard Industrial Classification (SIC). These sectors are listed in Table 1, along with the Major Group categories that they include, their consumption weights (ξ^j), the estimates of their production function parameters (that is, ν^j, α^j, and γ^j) and the estimate of their price rigidity parameters (ϕ^j).

Table 2 contains the capital adjustment cost parameter (χ) and the parameters of the shock processes. Note that shock heterogeneity is limited to the Division level of the SIC. That is, the authors assumed one distribution each for agriculture (Division A), all mining sectors (Division B), construction (Division C), all manufacturing sectors (Division D), and all services sectors (Divisions E through I). However, since draws are independent across sectors, shock realizations will be different in different sectors, whether they are in the same Division or not.

The discount rate (β), depreciations rate (δ), and elasticity of substitution between goods produced in the same sector (θ) were set to 0.997, 0.02, and 8, which are standard values in the literature. Finally, the input weights ζ_{ij} and κ_{ij} were computed as the share of sector i in the materials and investment input expenditures by sector j, respectively, using data from the Use Table and Capital Flow Table of the 1992 U.S. Input-Output (I-O) accounts.

Table 1. Model Parameters

Sector	SIC Codes	Consumption Weights	Production Function ν^j	α^j	γ^j	Price Rigidity
Agriculture	01 – 09	0.02	0.261	0.142	0.597	0.001
Metal Mining	10	0.01	0.328	0.306	0.366	4.81
Coal Mining	12	0.01	0.432	0.194	0.374	2.80
Oil and Gas Extraction	13	0.01	0.176	0.456	0.368	0.056
Nonmetallic Mining	14	0.01	0.314	0.254	0.432	81.42
Construction	15 – 17	0.01	0.394	0.052	0.554	140.7
Food Products	20	0.12	0.161	0.084	0.755	189.9
Tobacco Products	21	0.01	0.146	0.290	0.564	0.001
Textile Mill Products	22	0.01	0.229	0.067	0.704	13.78
Apparel	23	0.04	0.325	0.060	0.615	666.7
Lumber and Wood	24	0.01	0.247	0.100	0.653	70.88
Furniture and Fixtures	25	0.02	0.365	0.079	0.557	158.3
Paper	26	0.02	0.261	0.136	0.603	1.46
Printing and Publishing	27	0.01	0.398	0.124	0.478	24.72
Chemicals	28	0.03	0.237	0.183	0.581	0.199
Oil Refining	29	0.03	0.091	0.103	0.806	1.80
Rubber and Plastics	30	0.01	0.323	0.091	0.586	4.79
Leather	31	0.01	0.326	0.089	0.585	330.7
Stone, Clay and Glass	32	0.01	0.369	0.125	0.507	21.33
Primary Metal	33	0.01	0.229	0.084	0.687	507.5
Fabricated Metal	34	0.01	0.346	0.104	0.549	0.009
Nonelectric Machinery	35	0.01	0.361	0.112	0.527	0.001
Electric Machinery	36	0.02	0.350	0.127	0.523	0.005
Transportation Equip.	37	0.05	0.283	0.080	0.637	42.75
Instruments	38	0.01	0.460	0.100	0.440	0.001
Misc. Manufacturing	39	0.01	0.327	0.117	0.555	4.29
Transport and Utilities	40 – 49	0.21	0.314	0.248	0.437	151.1
Trade	50 – 59	0.25	0.500	0.148	0.352	423.8
FIRE	60 – 67	0.01	0.283	0.356	0.361	0.004
Other Services	70 – 87	0.01	0.427	0.195	0.378	0.305

Note: FIRE stands for finance, insurance and real estate. The consumption weights are based on Horvath (2000, p. 87). The production function parameters were estimated using the KLEM data set collected by Dale Jorgenson. Sectoral price rigidities were estimated by the Simulated Method of Moments. See Bouakez, Cardia and Ruge-Murcia (2009) for additional details.

Table 2. Other Model Parameters

Description	
Capital adjustment parameter	4.710
AR coefficient of productivity shock	
Agriculture	0.922
All mining sectors	0.827
Construction	0.852
All manufacturing sectors	0.949
All service sectors	0.763
SD of productivity innovation	
Agriculture	0.111
All mining sectors	0.063
Construction	0.024
All manufacturing sectors	0.033
All service sectors	0.020
AR coefficient of labor supply shock	0.984
SD of labor supply innovation	0.012
AR coefficient of money demand shock	0.711
SD of money demand innovation	0.186
AR coefficient of monetary policy shock	0.456
SD of monetary policy innovation	0.008

Note: Taken from Table 7 in Bouakez, Cardia and Ruge-Murcia (2009).

References

[1] Bagliano F.C. and Morana C., (2009), "International Macroeconomic Dynamics: A Factor Vector Autoregressive Approach," *Economic Modelling* 26 (2), pp 432-444

[2] Bai, J. (2003) "Inferential Theory for Factor Models of Large Dimensions", *Econometrica* 71, 135-171.

[3] Bai, J., and Ng, S., (2007), "Determining the Number of Primitive Shocks in Factor Models," *Journal of Business and Economic Statistics* 25, pp. 52-60.

[4] Bai, J., and S. Ng (2006), "Confidence Intervals for Diffusion Index Forecasts and Inference for Factor Augmented Regressions," Econometrica 74, pp. 1133-1150

[5] Bernanke, B.S., Boivin, J., and Eliasz, P., (2005), "Measuring Monetary Policy: A Factor Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics* 120, pp. 387-422

[6] Blanchard, O. J. and Khan, C. M. (1980), "The Solution of Linear Difference Models Under Rational Expectations," *Econometrica* 48, pp. 1305-1311.

[7] Boivin, J., Giannoni, M., and Mihov, I., (2009), "Sticky Prices and Monetary Policy: Evidence from Disaggregated U.S. Data," *American Economic Review* 99, pp. 350–84.

[8] Boivin, J., Giannoni, M., and Mojon, B., (2008), "How Has the Euro Changed the Monetary Transmission?," *NBER Macroeconomic Annual*

[9] Bouakez, H., Cardia E., and Ruge-Murcia, F. J., (2009), "Sectoral Price Rigidity and Aggregate Dynamics," CIREQ Working Paper 01-2009.

[10] Cattell, R. B., (1966), "The Scree Test for the Number of Factors," *Multivariate Behavioral Research* 1, pp. 245-76.

[11] Chamberlain, G., and Rothschild, M., (1983), "Arbitrage, Factor Structure, and Mean-Variance Analysis on Large Asset Markets," *Econometrica* 51, pp.1281-1304.

[12] Eickmeier, S. and Breitung, J., (2006), "How Synchronized are New EU Member States with the Euro Area? Evidence from a Structural Factor Model," *Journal of Comparative Economics* 34, pp. 538-563.

[13] Fernandez-Villaverde, J., Rubio-Ramirez, J. F., Sargent, T., and Watson, M. W., (2007), "ABCs (and Ds) of understanding VARs," *American Economic Review* 97, pp. 1021–1026.

[14] Foester, A.T., Sarte, P.-D., and Watson, M. W., (2008), "Sectoral *vs.* Aggregate Shocks: A Structural Factor Analysis of Industrial Production," Princeton University, Manuscript.

[15] Forni, M., Hallin, M., Lippi, M., and Reichlin, L., (2000), "The Generalized Dynamic-Factor Model: Identification and Estimation," *Review of Economics and Statistics* 82, pp 540-554.

[16] Forni, M., Hallin, M., Lippi, M., and Reichlin, L., (2000), "The Generalized Dynamic-Factor Model: One Sided Estimation and Forecasting," *Journal of the American Statistical Association* 100, pp. 830-840.

[17] Forni, M., and Gambetti, L., (2009), "The Dynamic Effects of Monetary Policy: A Structural Factor Model Approach," Manuscript.

[18] Forni, M., Giannone, D., Lippi, M., and Reichlin, L., (2009), "Opening the Black Box: Structural Factor Models with Large Cross Sections," *Econometric Theory* 25, pp. 1319-1347.

[19] Giannone D., and Reichlin, L., (2006), "Does Information Help Recovering Structural Shocks from Past Observations?" *Journal of the European Economic Association* 4, pp. 455-465.

[20] Gilchrist, S., Yankov, V., and Zakrajsek E., (2009), "Credit Market Shocks and Economic Fluctuations: Evidence from Corporate Bond and Stock Markets," *Journal of Monetary Economics* 56, pp. 471-493.

[21] Hallin, M., and Liska R., (2007), "The Generalized Dynamic Factor Model: Determining the Number of Factors," *Journal of the American Statistical Association* 102, pp. 603-617.

[22] Lescaroux, F., and Mignon, V., (2009), "Measuring the Effects of Oil Prices on China's Economy. A Factor Augmented Vector Autoregressive Approach," *Pacific Economic Review* 14, pp. 410-425.

[23] Lippi, M. and Reichlin, L., (1994) "VAR Analysis, Non-Fundamental Representations, Blaschke matrices", Journal of Econometrics 63, pp. 307-325.

[24] Moench, E. (2008), "Forecasting the Yield Curve in a Data-Rich Environment: A No-Arbitrage Factor-Augmented VAR Approach," *Journal of Econometrics* 146, pp. 26-43

[25] Mumtaz H. and Surico P., (2009), "The Transmission of International Shocks: A Factor-Augmented VAR Approach," *Journal of Money Credit and Banking* 41, pp. 71-100

[26] Onatski, A., (2009), "Testing Hypotheses About the Number of Factors in Large Factor Models," *Econometrica*, forthcoming.

[27] Onatski, A., (2005), "Asymptotics of the Principal Components Estimator of Large Factor Models with Weak Factors," Columbia University, Manuscript.

[28] Stock, J., and Watson, M. W., (2002), "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics* 20, pp. 147-162.